

**T.C.
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ÇOKLU ORTAM SES DOSYALARININ ZAMAN-FREKANS YÖNTEMİYLE
ANALİZİ VE KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

Hakan UYANIK

Elektrik Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Murat KÖSEOĞLU

TEMMUZ 2021

**T.C
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ÇOKLU ORTAM SES DOSYALARININ ZAMAN-FREKANS YÖNTEMİYLE
ANALİZİ VE KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

**Hakan UYANIK
36183615001**

Elektrik Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Murat KÖSEOĞLU

TEMMUZ 2021

TEŞEKKÜR VE ÖNSÖZ

Çalışmalarım boyunca maddi manevi destekleriyle beni hiçbir zaman yalnız bırakmayan aileme,

Bu tez çalışmasının her aşamasında yardım, öneri, bilgi, tecrübe ve desteklerini esirgemedi beni her konuda yönlendiren danışman hocam Sayın Dr. Öğr. Üyesi Murat KÖSEOĞLU'na,

Lisans ve lisansüstü öğretimim boyunca tecrübe ve bilgileriyle bana her konuda destek olan İnönü Üniversitesi Elektrik Elektronik Mühendisliği Bölümü'ndeki tüm hocalarıma,
teşekkür ederim.



ONUR SÖZÜ

Yüksek lisans tezi olarak sunduğum “Çoklu Ortam Ses Dosyalarının Zaman-Frekans Yöntemiyle Analizi ve Karşılaştırılması” başlıklı bu çalışmanın bilimsel ahlâk ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığına ve yararlandığım bütün kaynakların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Hakan Uyanık



İÇİNDEKİLER

TEŞEKKÜR VE ÖNSÖZ	i
ONUR SÖZÜ.....	ii
İÇİNDEKİLER.....	iii
ÇİZELGELER DİZİNİ.....	iv
ŞEKİLLER DİZİNİ.....	v
SEMBOLLER VE KISALTMALAR	vii
ÖZET.....	viii
ABSTRACT.....	ix
1. GİRİŞ.....	1
1.1 Genel Bilgiler.....	1
1.2 Tezin Amacı.....	3
2. TEORİK TEMELLER VE KAYNAK ÖZETLERİ	4
2.1 Literatür	4
2.1.1 Konuşma sinyallerinin analizi	4
2.1.2 Çevresel seslerin analizi	8
2.1.3 Müzik sinyallerinin analizi	9
2.2 Ses Sinyali.....	11
2.3 Fourier Analizi	12
2.3.1 Ayrık zamanlı Fourier dönüşümü.....	13
2.3.2 Hızlı Fourier dönüşümü.....	14
2.4 Pencere Fonksiyonları.....	21
2.4.1 Flat-top penceresi ve özellikleri	22
2.4.2 Hanning penceresi ve özellikleri	23
2.4.3 Dikdörtgen pencere ve özellikleri.....	24
2.5 Spektrogram.....	27
2.6 Ses Parmak İzi	32
2.6.1 Philips tekniği	39
2.6.2 Shazam tekniği	41
3. METODOLOJİ.....	42
3.1 Ses Parmak İzlerinin Üretilmesi	42
3.2 Ses Parmak İzlerinin Karşılaştırılması.....	51
4. ARAŞTIRMA BULGULARI VE TARTIŞMALAR.....	53
4.1 Eşik Değerli Spektral Nokta Seçiminin Gürültü Gürbüzlüğü	53
4.2 Farklı Pencere Fonksiyonlarının Algoritmanın Performansı Üzerindeki.....	56
Etkileri	56
4.3 Bazı Spektrogram Parametrelerinin Etkilerinin Araştırılması.....	59
5. SONUÇ VE ÖNERİLER	78
KAYNAKLAR.....	79
ÖZGEÇMİŞ	89

ÇİZELGELER DİZİNİ

Çizelge 2.1 : Bazı yaygın kullanılan Matlab pencere fonksiyonları.	22
Çizelge 4.1 : Veri tabanındaki çoklu ortam ses dosyalarına ait özellikler.	53
Çizelge 4.2 : Çeşitli gürültü oranlarına karşılık oluşan hash değeri veri sayıları.....	54
Çizelge 4.3 : Gürültü seviyelerine karşı elde edilen eşleşme sayıları.	55
Çizelge 4.4 : Test edilen ses dosyalarının müzik türleri.	59
Çizelge 4.5 : Farklı spektrogram parametrelerine göre 10 farklı ses dosyası için 10 saniyelik bölümlerine ait farklı gürültü tipleri ile yapılan testlere ait ... eşleşme sayılarının ve tanıma sürelerinin ortalamaları.	67
Çizelge 4.6 : Gürültüsüz durumda 512;%50;512 spektrogram parametre..... kombinasyonuna göre on farklı ses dosyasının 10 saniyelik eşleşme... sayılarını gösteren karışıklık matrisi (confusion matrix).	76



ŞEKİLLER DİZİNİ

Şekil 2.1	: 20. yüzyıl başlarında bir transkripsiyon havuzu.Kadınlar silindir dikte makineleri kullanıyor ve kayıtları kulak tüpleriyle dinliyorlar [7, 10]....	6
Şekil 2.2	: Rakam tanıyıcı devrelere ait blok şematığı [7, 11].....	7
Şekil 2.3	: Rakamların formant gösterimi [7, 11].	8
Şekil 2.4	: Bir ses sinyali örneği.	11
Şekil 2.5	: Bir kare dalganın sinüzoidal sinyaller ile ifadesi [56].	13
Şekil 2.6	: Cooley ve Tukey tarafından oluşturulmuş tek aşamalı 8 birim uzunlukta Radix-2 FFT'nin devre şeması [56].	15
Şekil 2.7	: 8192 örnekten oluşan ayrık zamanlı bir sinyale ait FFT grafiği.	16
Şekil 2.8	: 220, 440, 500, 1000, 1300, 1800, 2100, 2500,2700, 3200, 3500, 3700 Hz'lik sinüzodial sinyallerin toplamından oluşan bir sinyale ait..... spektrum grafiği.	17
Şekil 2.9	: 1024 örnek uzunluğundaki bir ses sinyaline ait FFT dönüşümüne ait... PSD çıktısı.	18
Şekil 2.10	: Matlab <code>freqz()</code> fonksiyonu yardımıyla bir sinyale ait FFT genlik ve FFT faz çıktısı [56].	19
Şekil 2.11	: Bir sinyale ait <code>periodogram()</code> fonksiyonu yardımıyla hesaplanmış FFT genlik verisi [56].	20
Şekil 2.12	: 64 örnek uzunlukta Flattop pencere fonksiyonunun zaman-frekans..... domeninde gösterimi.	23
Şekil 2.13	: 64 örnek uzunlukta Hanning pencere fonksiyonunun zaman-frekans... domeninde gösterimi.	24
Şekil 2.14	: 64 örnek uzunlukta dikdörtgen pencerenin zaman-frekans domeninde gösterimi.	25
Şekil 2.15	: Pencereleme işleminin spektruma etkisi [56].	26
Şekil 2.16	: Pencereleme işleminin spektruma etkisi [56].	26
Şekil 2.17	: Bir ses sinyaline ait spektrogram örneği.	29
Şekil 2.18	: Bir ses sinyaline ait spektrogram örneği. (a) 2-boyutlu görünümü, (b).. 3-boyutlu görünümü.	30
Şekil 2.19	: Zaman ve frekans çözünürlüğü arasındaki ilişki [76].	31
Şekil 2.20	: Ses parmak izi oluşturma şeması [1].	35
Şekil 2.21	: Philips ses parmak izi sistemine ait blok diyagram.	39
Şekil 2.22	: Philips ses parmak izi algoritması ile oluşturulmuş (a) Orijinal müziğe ait oluşturulmuş hash değerleri, (b) sıkıştırılmış ses dosyasına ait..... üretilmiş hash değerleri, (c) a ve b şekilleri arasındaki bit hataları..... (Siyah bölgeler) (BER=0.115-Bit Error Rate) [79].	40
Şekil 3.1	: Ses parmak izi üretimine ait çalışmada izlenen yöntemin blok diyagramı.	42
Şekil 3.2	: İki kanallı ses verisinden tek kanallı ses verisinin elde edilmesi. Yatay eksen zamanı, düşey eksen ise genliği ifade etmektedir.	43
Şekil 3.3	: Veri tabanından rastgele seçilmiş bir ses dosyasının 10 saniyelik bir.... bölümüne ait spektrogram görseli	44
Şekil 3.4	: Alt örnekleme işlemi.	45
Şekil 3.5	: Bir ses sinyaline ait spektrogram örneği.	46
Şekil 3.6	: Zaman-frekans çözünürlüğünün spektrogram üzerindeki etkisi [139]...	47

Şekil 3.7	: (a) Hedef bölge şeması. (b)-(c) Seçkin noktaların eşleştirilmesi (Landmarks) (d) Üretilen ses parmak izi verilerinin kaydedildiği tablo yapısı. Soldaki tablo ses parmak izi veri tabanına, sağdaki ise sorgusu yapılan ses verisine ait oluşturulan ses parmak izi verilerinin..... kaydedildiği tablodur. Sorgulama işlemleri bu iki tablo arasında gerçekleştirilir.....	50
Şekil 3.8	: Geleneksel karşılaştırma tekniği [142].....	52
Şekil 4.1	: Uygulamaya ait akış diyagramı.....	53
Şekil 4.2	: Üç farklı pencere fonksiyonu için farklı gürültü seviyelerine karşı elde edilen eşleşme sayıları.....	57
Şekil 4.3	: Veri tabanı içerisinde bulunan bir çoklu ortam ses dosyasının (Song... ID:18) 10 saniyelik bir kısmına uygulanan farklı tür ve seviyelerdeki gürültü ekleme işlemi sonucu elde edilen spektrogram çıktıları..... Kırmızı noktalar spektral tepeleri temsil etmektedir: (a) Gürültüsüz ... durum, (b) 3 dB AWGN, (c) 4.8 dB restoran ambiyans gürültüsü, (d) 13.5 dB trafik ambiyans gürültüsü.	60
Şekil 4.4	: Farklı iki ses dosyasına ait AWGN gürültüsü ile gerçekleştirilen FFT sayısı parametresine ait test sonuçları.(a) 26 nolu dosya için sonuçlar (b) 42 nolu dosya için sonuçlar.	62
Şekil 4.5	: Veri tabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç farklı gürültü tipi ile gerçekleştirilen FFT sayısı etkisine ait sonuçlar. (a) AWGN tipi gürültü için sonuçlar, (b) Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar... 63	63
Şekil 4.6	: Veri tabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç..... farklı gürültü tipi ile gerçekleştirilen üstüste binme oranı etkisine ait sonuçlar.(a) AWGN tipi gürültü için sonuçlar, (b) Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar.... 65	65
Şekil 4.7	: Veri tabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç farklı gürültü tipi ile gerçekleştirilen pencere uzunluğu etkisine ait sonuçlar (a)AWGN tipi gürültü için sonuçlar, (b)Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar..... 66	66
Şekil 4.8	: 18 nolu şarkının rastgele seçilen 10 saniyelik bölümü için oluşan test.. sonuçları. 5 farklı spektrogram parametre kombinasyonuna göre elde edilen değerler yatay eksen de dosya numarasını, düşey eksen de ise.... eşleşme sayılarını gösterecek şekilde ifade edilmiştir..... 69	69
Şekil 4.9	: AWGN gürültüsü için artan SNR seviyelerine karşı (a) pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki.... değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri .. için elde edilen ortalama NRR değerlerinin değişimine ait grafikler... 71	71
Şekil 4.10	: Restoran gürültüsü için artan SNR seviyelerine karşı (a) pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki ... değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri için elde edilen ortalama NRR değerlerinin değişimine ait grafikler.... 73	73
Şekil 4.11	: Trafik gürültüsü için artan SNR seviyelerine karşı (a)pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri için elde ... edilen ortalama NRR değerlerinin değişimine ait grafikler 75	75

SEMBOLLER VE KISALTMALAR

FT	: Fourier dönüşüm
HAS	: İnsan duyu sistemi
MASK	: Maskelenmiş ses spektral anahtar noktaları
CBID	: İçerik tabanlı ses tanımlama
DCT	: Ayırık kosinüs dönüşümü
MCLT	: Modüle edilmiş karmaşık dönüşüm
SUI	: Konuşma kullanıcı arayüzü
MUI	: Çok modlu kullanıcı arayüzü
IIR	: Sonsuz dürtü yanıtı
FIR	: Sonlu dürtü yanıtı
STFT	: Kısa-süreli Fourier dönüşümü
ISTFT	: Ters kısa-süreli Fourier dönüşümü
SIFT	: Ölçekle değişmeyen özellik dönüşümü
IFPI	: Uluslararası fonografik endüstrisi federasyonu
RIAA	: Amerikan kayıt endüstrisi birliği
RMS	: Kök kare ortalama
AWGN	: Toplamsal beyaz Gauss gürültüsü
SNR	: Sinyal gürültü oranı
DFT	: Ayırık zamanlı Fourier dönüşümü
IDFT	: Ters ayırık zamanlı Fourier dönüşümü
PSD	: Spektral güç yoğunluğu

ÖZET

Yüksek Lisans Tezi

ÇOKLU ORTAM SES DOSYALARININ ZAMAN-FREKANS YÖNTEMİYLE ANALİZİ VE KARŞILAŞTIRILMASI

HAKAN UYANIK

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı

89+VIII sayfa

2021

Danışman: Dr. Öğr. Üyesi Murat KÖSEOĞLU

Ses sinyali arama algoritmaları, özellikle çeşitli multimedya uygulamalarında ve akıllı cihazlarda büyük veri tabanlarında gerekli olan dosyayı tespit etmek için sıklıkla kullanılırlar. Genellikle farklı ses parmak izi çıkarma tekniklerine dayanırlar ve dijital dünyada gelişen teknolojiyle birlikte, özellikle son yirmi yılda çok popüler hale gelmişlerdir. Ayrıca, bu teknikten esinlenen benzer algoritmalar farklı teknolojik alanlarda uygulanmış ve oldukça başarılı sonuçlar ortaya koymuşlardır. Farklı ses parmak izi çıkarma yöntemlerine dayalı olarak geliştirilen bu algoritmaların, verilen görevi mümkün olan en kısa sürede güvenilir ve sağlam bir şekilde gerçekleştirmesi öncelikli hedeftir.

Bu çalışmada, literatürde izlenen geleneksel ses sinyali arama metodolojisi kullanılarak veri tabanında doğru eşleşmeyi tespit etmek için, spektral tepeler yöntemine dayalı mevcut arama algoritmalarına alternatif bir ses parmak izi temelli ses arama programı oluşturulmuştur. Programın doğru eşleştirmedeki başarısının gösterilmesinden sonra, program içerisinde kullanılan pencere fonksiyonunun programın performansı üzerindeki etkisi ve yine program boyunca önemli bir yer tutan spektrogram yapısındaki bazı parametrelerin programın güvenilirliği ve gürbüzlüğü üzerindeki etkileri test edilmiştir. Böylelikle benzer metodoloji kullanılarak analiz edilen ses dosyalarının tanınmasında, bahsedilen faktörlerin etkisinin araştırılarak spektral tepeler yöntemine dayalı yeni çalışmalara katkı sağlanması hedeflenmiştir. Yapılan çalışmalar sonucunda sorgusu yapılan ses dosyasının, her koşulda tatmin edici bir güvenilirlik ve sağlamlıkla veri tabanında tespit edildiği gözlemlenmiştir.

Anahtar Kelimeler: Ses sinyali, Ses parmak izi, Spektral tepeler metodu

ABSTRACT

Master Thesis

COMPARISON AND ANALYSIS OF MULTIMEDIA AUDIO FILES WITH TIME-FREQUENCY METHOD

HAKAN UYANIK

Inonu University
Graduate School of Nature and Applied Sciences
Department of Electrical and Electronic Engineering

89+VIII pages

2021

Supervisor: Assist. Prof. Dr. Murat KÖSEOĞLU

Audio signal searching algorithms are often used to detect the required file in large databases, especially in various multimedia applications and smart devices. They generally rely on different audio fingerprinting techniques and have become very popular especially in the last two decades with the advancing technology in the digital world. In addition, similar algorithms inspired by this technique have been applied in different technological fields and have produced very successful results. The primary goal of these algorithms, which are developed based on different audio fingerprinting methods, is to perform the given task in the shortest possible time in a reliable and robust manner.

In this study, using the traditional audio signal search methodology followed in the literature, an alternative audio fingerprint-based audio search program to the existing search algorithms has been created based on the spectral peaks method in order to determine the correct match in the database. After demonstrating the success of the program in correct matching, the effect of the window function used in the program on the performance of the program and the effects of some parameters in the spectrogram structure, which has an important role throughout the program, on the reliability and robustness of the program were tested. Thus, it is aimed to contribute to new studies based on the spectral peaks method by investigating the effect of the mentioned factors in the recognition of audio files analyzed using a similar methodology. As a result of the studies carried out, it has been observed that the audio file that was queried was detected in the database with a satisfactory reliability and robustness under all conditions.

Keywords: Audio signal, Audio fingerprint, Spectral peaks method

1. GİRİŞ

1.1 Genel Bilgiler

Günümüzde sinyal işlemenin önemi her geçen gün artmakta, çeşitli sinyal işleme teknikleri ve algoritmaları geliştirilmektedir. Sinyal işleme yöntemleri biyomedikal, tıp, güvenlik, haberleşme, robotik, yer bilimleri, otomotiv, medya ve iletişim gibi pek çok önemli alanda, farklı sinyal türlerine uygulanmaktadır.

20. yüzyılın ortalarından itibaren, yarı-iletken endüstrisinin doğuşuyla birlikte, dijital elektronik alanında yaşanan gelişmeler neticesinde sinyal işleme alanında yeni bir dönem başlamış oldu. Sinyaller üzerinde göstermiş olduğu işlem yeteneği sayesinde, dijital elektronik yardımıyla, sinyallere ait özellikler hakkında bilgi edinimi ve elde edilen bu bilgiler ile sinyallerin isteğe göre düzenlenmesi imkânlarına erişilmiştir.

Ses sinyallerinin analizi, genellikle kişisel ses tanıma, çoklu ortam ses dosyalarının yasal telif haklarının korunması, ekoloji, müzik endüstrisi, adli saha konuları, akıllı uygulamalarda kullanılan sesli kontrol sistemleri ve tıp alanında tanı koyma gibi daha birçok alanda kendisine yer edinmiştir. Aynı zamanda, müzik parçalarının 3-4 saniye gibi kısa örnekleri ile tanımlanmasını sağlayan Shazam ve SoundHound gibi uygulamalar temelinde ses işleme algoritmaları kullanılmaktadır.

Ses sinyalleri analiz edilirken, genellikle Fourier dönüşüm (Fourier Transform-FT) kullanılarak frekans domenine aktarılmış sinyalin, önceden belirlenmiş ayırt edici birtakım parametrelerinin ölçümleri yapılır. Bu aşamada çok çeşitli yöntemler kullanılabilir. Buradaki amaç boyutluluğu azaltmak ve sesteki bozulmalara karşı dayanıklılığı arttırmaktır. Algısal olarak daha anlamlı öznelilikler çıkarabilmek için insan duyu sistemiyle (Human Auditory System-HAS) ilgili bilgilerin göz önünde bulundurulması çokça rastlanır bir durumdur.

Ses sinyallerinin analog ve dijital olarak işlenmesi, mekanik kayıt cihazlarının ve telefonun ilk kullanıldığı zamanlardan günümüze kadar bir asırdan daha uzun bir geçmişe sahiptir. Günümüz bilgi çağında, ses dosyası oluşturma ve depolama işlemleri, geleneksel yöntemlerden sıyrılıp, dijital çağa uyum sağlamıştır. İnternet ve çoklu ortam teknolojilerinin hızlı gelişimine bağlı olarak, ses dosyaları da eskiye nazaran daha büyük boyutlarda, çeşitli

formatlarda ve daha yüksek çözünürlüklerde üretilmektedir. Bu durum bu dosyaların analizinde yeni zorlukları beraberinde getirmektedir. Artan veri miktarları, veri türleri ve çözünürlük seviyeleri bu dosyaların analizinde yeni yöntemlere ihtiyaç duyulmasına yol açmaktadır.

Ses sinyallerinin ayırt edilmesi, farklı birçok alanda çözülmesi gereken bir problemdir. Ses parmak izi teknolojisi, bu problemi çözmede kullanılan ve oldukça başarılı sonuçlar veren, kendisini kanıtlamış bir yöntemdir.

Ses parmak izi, insanlardaki parmak izinde olduğu gibi, ses dosyalarına ait eşsiz bir kimlik oluşturarak bu dosyaların ayırt edilmesine olanak sağlayan bir metottur. Bu teknoloji sayesinde üzerinde işlem yapılan verinin, ham haldeki tüm bileşenlerinin işlenmesi yerine, bu veriye nazaran oldukça küçük boyutlarda kalan, bu dosyadan türetilmiş parmak izlerinin işlenmesi yeterli olmaktadır. Bu yöntem sayesinde veri kullanımı ve depolama işlemleri ile işlem süreleri asgâri düzeye çekilmektedir.

Ses parmak izi yöntemi, kendisinden daha önce oluşturulan filigranlama (watermarking) yönteminden birçok yönden farklıdır. Filigran yapısı dosya içine gömülüdür ve dosyaya ait verilerde değişikliklere yol açar ve genel olarak telif hakları, korsanlık takibi ve sertifikalandırma işlemleri gibi uygulamalarda tercih edilir. Ses parmak izi ise ses dosyasına ait veriye hiçbir şekilde dahil olmaz. Dosyanın algısal yapısını temel alarak işlemlerini gerçekleştirir.

Ses parmak izi tekniğinin bazı yaygın kullanım alanları; içerik tabanlı ses tanımlama (Content-Based Audio Identification-CBID), içerik tabanlı bütünlük doğrulaması (Content-Based Integrity Verification) ve filigranlama uygulamaları (Watermarking Support) şeklinde sıralanabilir [1].

Philips Technique [2], the Shazam Algorithm [3] ve Google Waveprint [4] ses parmak izi algoritmaları, ticari alanda büyük başarılarla ulaşmış en önemli yaklaşımlardır. Bu yaklaşımların dışında, yakın zamanda birçok yeni yöntem ortaya çıkmış ve çıkmaya da devam etmektedir. Bunlar arasında öne çıkan bir başka çalışma Anguera ve arkadaşları tarafından tasarlanan MASK (Masked Audio Spectral Keypoints–Maskelenmiş Ses Spektral Anahtar Noktaları) yöntemidir [5]. Bu yöntem henüz yukarıda belirtilen yöntemler kadar yaygın kullanım şansına ulaşamamış olsa da oldukça iyi sonuçlar vermektedir [6].

Yukarıda belirtilen yaklaşımlardan, özellikle ticari alanda kendisine yer bulmuş ve oldukça başarılı sonuçlara ulaşmış iki uygulama Philips ve Shazam algoritmalarıdır.

Bunlardan Philips algoritması, Birleşik Devletler'deki Gracenote isimli şirket ile Hollanda'lı şirket Royal Philips Research Institute'ün birlikte geliştirdikleri bir algortimadır. Shazam algoritması ise kendi adı ile kurulan Shazam Inc. şirketi çatısı altında oluşturulmuştur. Bu çalışmada bu algoritmalarından Shazam'a ait yöntemler üzerinde yoğunlaşmıştır.

1.2 Tezin Amacı

İnsanlar çevrelerinde duydukları bir müzik parçasını araştırırken, genel olarak o an hakkında bilgi sahibi oldukları noktalar üzerinden araştırma yapmaya çalışırlar. Bu bilgiler kimi zaman, eğer sanatçının ses tınısına bir aşinalık varsa sanatçı adı, kimi zaman da parça içerisinde geçen şarkı sözleri biliniyorsa, şarkı sözleri olabilmektedir. Bu tarz bilgilerle yapılan sorgulamalarda ise bir sonraki aşamada karşılırlarına başka kontrol etmeleri gereken noktalar çıkabilmekte ve çoğu zaman aradıkları veriye ulaşmaları birkaç aşama geçse dahi tamamlanamamaktadır. Eğer sorgusu yapılan parça, enstrümantal ise bu noktada yapabilecekleri pek bir şey kalmamaktadır. Bazende, duyulan müzik parçası, süre bazında çok kısa olabilmekte ve dolayısıyla algısal tespiti için gerekli bilgiler toplanamamaktadır. İşte bu noktada, son 20-30 yıldır, milyonlarca dijital ses dosyasının internet aracılığı ile herkesin kullanımına sunulmasıyla ve her geçen gün binlerce yenilerinin eklenmesiyle, ses dosyalarının miktarı çığ gibi büyümüş ve bu karmaşılaşan dünyada bu verilere hızlı ve güvenilir bir şekilde ulaşmak, bu alanda yeni bir ihtiyaç olarak belirmiştir.

Bu çalışmada çoklu ortam ses dosyalarının tanımlanmasına yönelik bir algoritma, daha önce bu alanda yapılan çalışmalar baz alınarak geliştirilmiş, ardından geliştirilen bu yöntem, çoklu ortam ses dosyalarına uygulanarak ve çeşitli testlere tâbi tutularak, geliştirilen yöntemin güvenilirliği, kararlılığı ve hızı konularında incelemeler ve değerlendirmeler yapılmıştır.

2. TEORİK TEMELLER VE KAYNAK ÖZETLERİ

2.1 Literatür

Literatüre bakıldığı zaman, ses verilerinin analiz işlemlerinin üç temel başlık altında yoğunlaştığı görülmektedir. Bunlar, konuşma sinyallerinin analizi, çevresel seslerin analizi ve müzik sinyallerinin analizi şeklindedir.

2.1.1 Konuşma sinyallerinin analizi

Konuşma sinyallerinin analizi, tarih boyunca en çok merak uyandıran sinyal analizi konularının başında gelmiştir. 1800'lü yılların başında Joseph Fourier'in kendi adıyla anılan Fourier analiz tekniğini bilim dünyasına sunması ile doğal sinyallerin harmonik sinüzoidlerinin spektrumunun elde edilmesi ile birlikte üzerinde en çok araştırma yapılan sinyal analizi konularından olmuştur [7]. Spektrum analizinin temelinde, bir fonksiyonun (sinyalin) farklı türden fonksiyonların kombinasyonu şeklinde ifade edilmesi ve böylelikle fiziksel olarak tasarlanabilmesi yatmaktadır. Burada sinyalin, trigonometrinin temel fonksiyonları olan sinüs ve kosinüs fonksiyonları türünden ifadesi temel hedeftir.

[7] nolu referansa göre, bu türden bir dönüşümle ilgili ilk matematiksel çalışmaları Leonhard Euler'in gerçekleştirdiği söylenebilir. Euler, 1748 yılında yayınlamış olduğu çalışmalarda, bazı fonksiyonların, sinüs ve kosinüs gibi sonsuz seriler türünden nasıl yeniden yazılabileceğini keşfetmiştir. Euler'in bu çalışmalarındaki hedefi bir frekans spektrumu analizi değil, bu seriler yardımıyla diferansiyel ifadeleri çözmektir.

Daniel Bernoulli, Euler'in bu düşüncesinin ilk uygulamalarından birini, 1753 yılında, titreşim fiziği alanında titreyen tel eşitliğinde kullanarak hayata geçirmiştir. Bu çalışmasından elde ettiği çözüm, doğrusal akustiğin temelini oluşturan dalga denklemi olarak adlandırılan, titreşen tel için diferansiyel denklemi gerçekten de ifade etmektedir [7].

Joseph Fourier 1807 yılında yayınlamış olduğu bir yazısında, ısı transferini yöneten diferansiyel denklemi çözmenin bir izlediği yolu, yani "keyfi" bir fonksiyon ile hem sinüs hem de kosinüs serisi temsilinin nasıl türetileceğini göstermiştir [7, 8]. Fourier ayrıca titreşen telin benzer bir diferansiyel denklem tarafından ifade edilebileceğinin farkındaydı. El yazısı notlarında, verilen yöntemin, soruna yönelik işlevsel çözümleri sunabilmek için de geçerli olacağını ve böylece dalga denkleminin çözümleri sağlayacağını belirtmiştir [7].

Fourier'in bu alandaki asıl ve önemli katkısı, tam olarak trigonometrik seri temsilinin icadı değil, daha ziyade bu tür bir temsilin genelliğini belirleme ve ayrı seriler formunu periyodik olmayan fonksiyonları temsil edebilen bir integral forma genişletme çalışmasıydı. Gerçekte, Fourier serisinin ve integralinin önemli bir kısmı, matematik alanında bir diğer tarihi dev olan Carl Gauss'un şahsi çabalarıyla bilim dünyasına kazandırılmıştır. Gauss'un günümüzde Fourier serisi ve integrali olarak adlandırılan el yazması toplu çalışmaları, 1866'da ortaya çıkana kadar yayınlanmamıştır. Tarihçiler bu özel el yazmasının büyük olasılıkla 1805'te yazıldığını kabul etmektedir [7, 9].

Tüm bu gelişmelerle birlikte, ses sinyali işleme alanında yeni bir kapı aralanmıştır. Yüz yıldan daha uzun süredir insanlığın ilgisini çeken bir başka konu da otomatik konuşma algılama sistemleri olmuştur. Bunun için yukarıda bahsedilen gelişmelerin ardından ses sinyallerinin kaydedilmesi öncelikli hedef olmuştur. 1881'de Alexander Graham Bell, kuzeni Chichester Bell ve Charles Sumner Tainter, üzerinde yukarı ve aşağı olukların bir kalemle kesilebildiği, balmumu kaplamalı dönen bir silindir kullanan bir kayıt cihazı icat ettiler [7, 10]. Bu cihaz, Bell'in daha önce telefonla kullanmak için icat ettiği ve gelen ses basıncına dayalı çalışan bir mikrofonla aynı şekilde çalışmaktaydı. Bu buluşun ardından, Bell ve Tainter, ofis ortamlarında ses kaydı ve yeniden üretimi için makineler üretmek amacıyla 1888'de Volta Graphophone Co.'yu kurdu. Daha sonra Columbia Graphophone Co. olarak isim değiştiren American Graphophone Co., 1907'de patenti aldı ve "Dictaphone" teriminin ticari markasını sahiplenmiş oldu. Tüm bu gelişmeler yaşanırken, Thomas Edison, fonograf icadına, bu tekniği balmumuna adapte edilen kalay folyo bazlı bir silindir kullanarak uygulamış ve icat ettiği bu buluşunu doğrudan Columbia ile rekabet etmek için "Ediphone" adı altında geliştirmiştir. Bu ürünlerin amacı, oluşturulan kayıtları çevrimdışı yazacak bir sekreter için (muhtemelen Şekil 2.1 de gösterildiği gibi bu hizmeti sunan büyük bir kayıt havuzunda) notların ve mektupların dikte edilmesini ve böylece maliyetli stenograflara olan ihtiyacı ortadan kaldırmaktı. On dokuzuncu yüzyılın başındaki bu "ofis mekanizasyon" kavramı, yirminci yüzyılın ortalarında ofis otomasyonunun çehresini değiştiren elektrikli daktilo da dahil olmak üzere bir dizi elektrikli ve elektronik aleti ve bir dizi iyileştirmeyi ortaya çıkardı. Bu noktada, balmumu silindirleri veya diğer kayıt ortamlarında konuşmanın kaydedilmesi ve işlenmesi sıkıntısıyla uğraşmak zorunda kalmadan bir insan sesine doğrudan yanıt verebilecek ve bunu kopyalayabilecek bir "otomatik daktilo" yaratma konusundaki merak bu alandaki bir sonraki hedef olmuştur [7, 10].

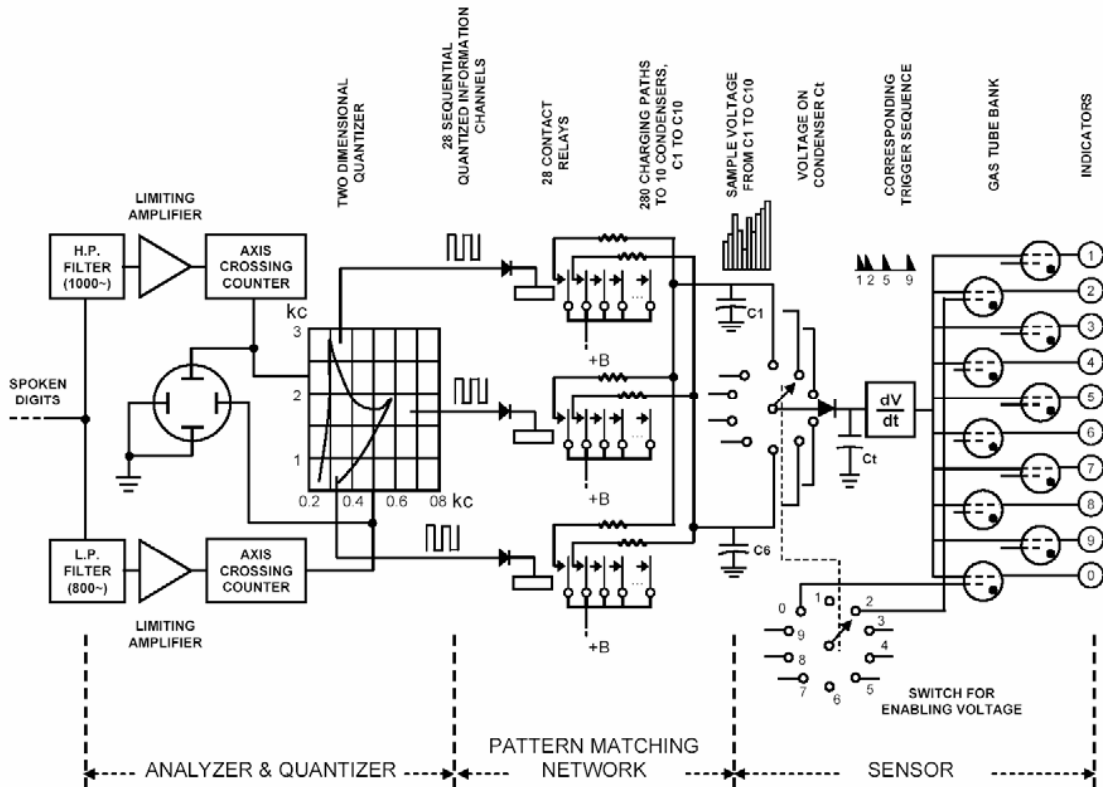


Şekil 2.1 : 20. yüzyıl başlarında bir transkripsiyon havuzu. Kadınlar silindir dikte makineleri kullanıyor ve kayıtları kulak tüpleriyle dinliyorlar [7, 10].

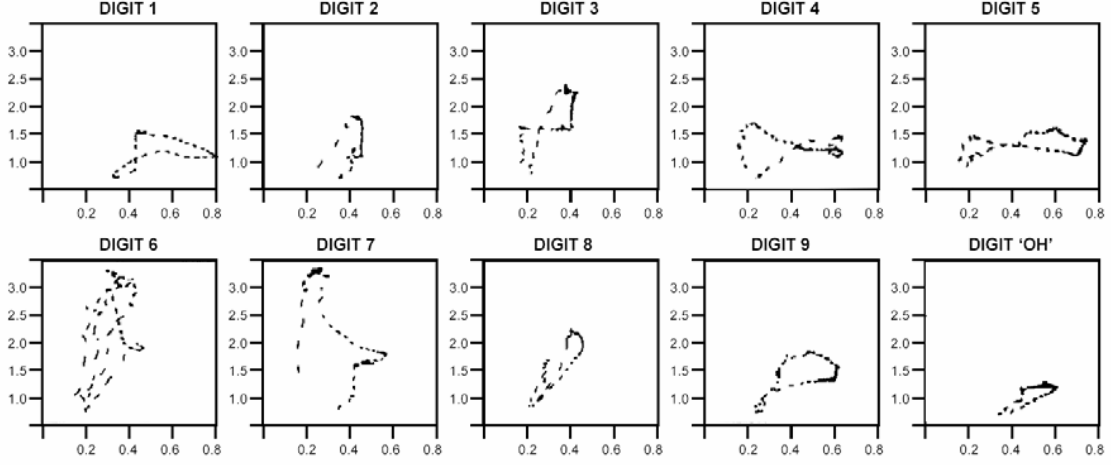
Konuşma tanıma teknolojisi, özellikle Stanley Kubrick'in beğeni toplayan filmi "2001: A Space Odyssey" olmak üzere 1960'ların ve 1970'lerin gişe rekorları kıran birçok filminde popüler hale gelmesi nedeniyle, dünya genelinde büyük ilgi duyulan bir konu olmuştur. Bu filmde, "HAL" adlı akıllı bir bilgisayar, doğal bir sesle konuşabiliyor ve akıcı bir şekilde konuşulan konuşmayı tanıyıp anlayarak buna göre yanıtlar verebiliyordu. HAL'ın bu antropomorfizmi, insanları makinelerin potansiyeli konusunda bilinçlendirmiştir. Ünlü Star Wars destanında George Lucas, akıllı makinelerin yeteneklerini hem mobil hem de akıllı hale getirerek genişletti ve böylece R2D2 ve C3PO gibi droidler doğal bir şekilde konuşabiliyor, akıcı konuşmayı tanıyıp anlayabiliyor ve çevrede dolaşp çevreleriyle etkileşim kurabiliyordu. Daha yakın zamanlarda (1988'de), teknoloji dünyasında Apple Computer, 2011 yılı için Konuşma Kullanıcı Arayüzü (Speech User Interface-SUI) ve Çok Modlu Kullanıcı Arayüzü (Multimodal User Interface-MUI) kavramlarını tanımlayan "Bilgi Gezgini" (Knowledge Navigator) adlı bir konuşma teknolojisi ve bilgisayar programı geliştirmiştir [7, 10].

Otomatik konuşma tanıma sistemleri tasarlamaya yönelik ilk girişimlere, çoğunlukla konuşmanın fonetik unsurlarını (dilın temel sesleri) tanımlayan ve sözlü bir ifadenin akustik olarak nasıl gerçekleştirildiğini açıklamaya çalışan akustik-fonetik teorisi esas alınarak başlanmıştır. Bu unsurlar, çeşitli fonetik bağlamlarda sesi üretmek için kullanılan fonemleri ve karşılık gelen eklem yeri (fonemlerin birleşim noktası) ve biçimlerini içerir. Örneğin, sabit bir ses üretmek için, ses tellerinin titreşmesi (ses yolunu uyarmak için) gerekir ve ses yolunda yayılan hava, bir akustikte meydana gelen benzer doğal rezonans modları ile sesi meydana getirir. Formantlar veya biçimlendirici frekanslar olarak adlandırılan bu doğal rezonans modları, konuşma gücü spektrumunda enerji yoğunlaşmasının ana bölgeleri olarak tezahür eder. 1952'de Davis, Biddulph ve Balashek, Bell Laboratuvarlarında, her bir rakamın sesli bölgelerinde ölçülen (veya tahmin edilen) biçimlendirme frekanslarını kullanarak tek bir konuşmacı için izole rakam tanıma sistemi geliştirmişlerdir [7, 11].

Şekil 2.2, Davis ve arkadaşları tarafından geliştirilen rakam tanıyıcıya ait blok diyagramını göstermektedir ve Şekil 2.3, sırasıyla on rakamın her biri için, birinci ve ikinci biçimlendirici frekansların boyutları boyunca biçimlendirici yörüngelerin çizimlerini göstermektedir. Bu yörüngeler, en iyi eşleşen basamak olarak, bilinmeyen bir rakam ifadesinin kimliğini belirlemek için "referans model" olarak kullanılmışlardır [7, 10].



Şekil 2.2 : Rakam tanıyıcı devrelere ait blok şematığı [7, 11].



Şekil 2.3 : Rakamların formant gösterimi [7, 11].

Günümüzde konuşma teknolojileri, oldukça farklı görevler için ticari olarak kullanılmaktadır. Bu teknolojiler, makinelerin insan sesine, doğru ve güvenilir bir şekilde yanıt vermesini sağlayarak, yararlı ve değerli hizmetler sunarlar.

2.1.2 Çevresel seslerin analizi

Çevresel sesler, tanım gereği oldukça geniş bir ses kaynağı gurubunu kapsamaktadır. Ekoloji, yer bilimleri gibi birçok önemli alana ait uygulamalarda kendisine yer bulmaktadır. En çok ön plana çıkan uygulama alanı ise bir kişinin günlük yaşamı hakkında fikir sahibi olmak için kullanılmasıdır. Bu alanda yararlanılabilecek zengin bir bilgi kaynağıdır. Günlük hayatımızdaki birçok aktivite, beraberinde bir ses verisi üretir (konuşmak, yürümek, el yıkamak gibi). Ayrıca, günlük hayatımızda bulunduğumuz çoğu yerin (trafik, restoran, stadyum, bar, market vb.) genellikle kendine has bir ses düzeni vardır. Bu ses farklılıklarına dayalı çeşitli “bağlam farkındalık” (context awareness) sistemleri geliştirilmiştir [12].

Sese dayalı bağlam farkındalık sistemlerinde, ortam türünün belirlenmesi, ortamda gerçekleşen etkinliklerin türlerinin belirlenmesi gibi bağlamsal bilgileri elde etmek için çevresel sesler sıklıkla kullanılmaktadır. [13-18] nolu çalışmalar bu alanda uygulanmış çalışmalara örnektir.

Ses tabanlı bağlam farkındalık sistemleri ayrıca robot navigasyonunda [19, 20], ses tabanlı gözetim sistemlerinde [21], ses tabanlı adli tıpta [22], işitme cihazlarında [23], yalnız yaşayan yaşlı insanlara yardımcı olmak için ev izleme ortamında [24, 25], akıllı evler [26] gibi birçok farklı alanda kullanılmaktadır [12].

2.1.3 Müzik sinyallerinin analizi

Müzik sinyallerinin analizi ilk olarak, ses analizinde kullanılan tekniklerin müzik ses sinyallerinin özelliklerine göre yorumlanmasıyla gerçekleştirilmiştir. Ancak bu tür sinyaller konuşma sinyallerinden farklı olarak, belirli akustik ve yapısal özelliklere sahiptirler. Bu nedenle bu tür sinyaller incelenirken armoni, ritim, akor, perde ve tını gibi daha birçok müzikal boyutun özel olarak incelenmesi büyük önem arz etmektedir.

Sinyal işlemenin müzik sinyallerine uygulanmasının 1920'lerin başında gerçekleştiği söylenebilir [27]. Bu konuda ilk örnek olarak, osilatörüne bağlı antenlerin kapasitansının bir kişi tarafından elle kontrolü ile çalınan theremin enstrümanı verilebilir. 1940'lı ve 1950'li yıllarda modern sinyal işlemenin gelişimi ile, Karlheinz Stockhausen gibi bestecilerin sinyal oluşturucuları, halka modülatörleri gibi sinyal üreticileri yardımıyla elektronik laboratuvarlarından doğrudan alınan sinyalleri kullanarak ilk elektronik müzik dalgasını başlatmışlardır. 1960'larda ve 1970'lerde kişisel dijital bilgisayarların ortaya çıkmasının ardından, Max Matthews ve John Pierce gibi öncüler tarafından müzik sentezlemek için kullanılmaları çok uzun sürmemiştir. Deneysel müzik, sinyal işlemenin yenilikçi uygulamalarının sabit bir kaynağı olarak kalmış ve sesleri sentezlemek ve değiştirmek için önemli bir karmaşık teknikler bütünü ortaya çıkarmıştır [27-29].

Perde, çokseslilik, zamanlama, vuruşlar gibi özellikler üzerinden analizi yapılan müzik sinyalleri, sinyal işleme alanında her dönem ilgi çekici bir alan olmuştur. Müzikte kroma özellikleri, tüm sinyale ait spektrumun, müzikal oktavın 12 farklı yarı tonunu (veya kroma) temsil eden 12 bölmeye yansıtıldığı, müzik sinyalleri hakkında bilgi edinmek için oluşturulmuş güçlü bir temsil şeklidir. Müzikte, bir oktav aralıklı notalar benzer olarak algılandığından, mutlak frekans (yani orijinal oktav) olmadan bile kroma dağılımını bilmek ses hakkında yararlı müzikal bilgiler verebilir ve hatta algılanan müzikal benzerliği ortaya çıkarabilir nitelikte olabilmektedir [27]. Bu durum orijinal spektrumlarda belirgin değildir [30]. [31-35] nolu çalışmalar kroma yapısı temel alınarak oluşturulmuş bazı önemli çalışmalardır.

Müzik sinyallerinde tonal nitelikler ve kroma özellikleri, enstrümantasyon türünden bağımsız olarak çok sesli (polifonik) müzik içeriğini yakalamanın daha önemli olduğu çeşitli müzik analizi ve sorgulama işlemleri için kullanılmıştır. Bu tür uygulamalar arasında akor tanıma, hizalama, "cover şarkı" algılama ve yapı analizi uygulamaları bulunur [27]. Bu

yapılardan başka, tempo, vuruş ve ritmin müzikal yönleri, müziğin anlaşılması ve müziğin etkileşimi için temel rol oynayan yapılardır [27, 36].

Müzik sinyallerinin analizinde kullanılan akor yapısı, en az üç perdeden elde edilen seslerin toplamından oluşan bir kavramdır [27, 37]. [38-46] nolu çalışmalar müzik sinyalleri içerisinde bu yapının incelenmesi ile oluşturulmuş bazı önemli çalışmalardır.

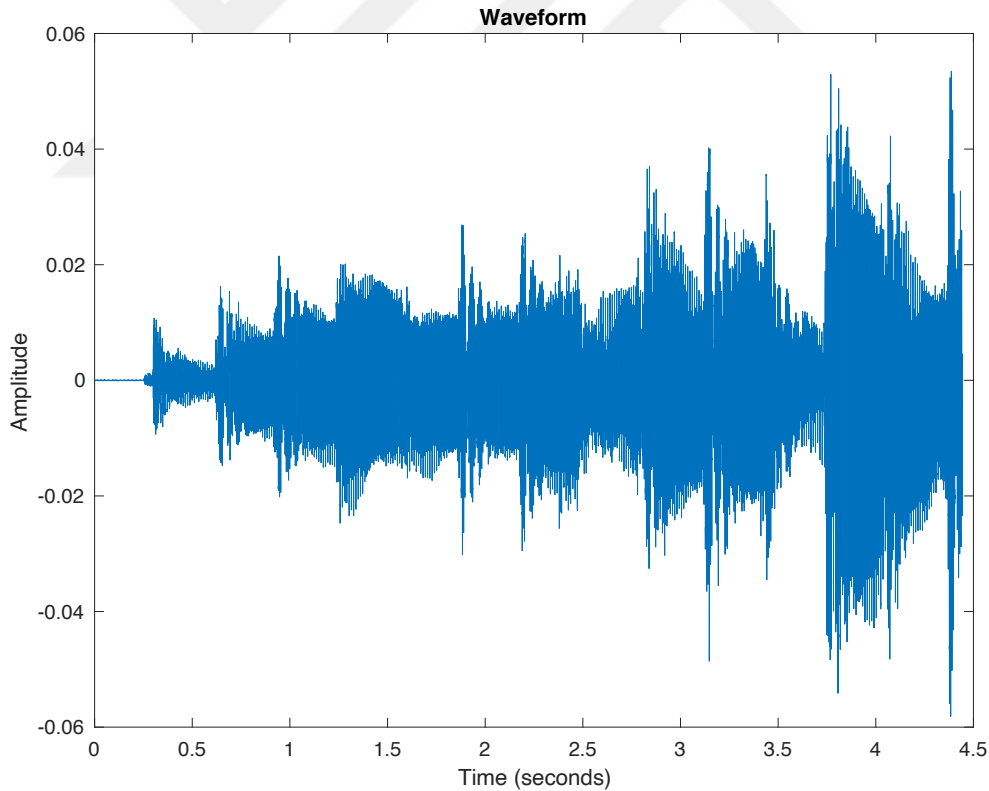
Kroma özellikleri, özellikle müzik sesinde akorları tanıma görevi için [38] nolu çalışmada “Perde Sınıf Profilleri” olarak tanımlanmıştır. Mevcut notaların ve içeriğin genel özelliği olarak, akorlar kısa bir pencerenin genel temsiline dayalı olarak tanınabilir. Dahası, notaları bir oktav kadar yukarı veya aşağı kaydırmanın akor kimliği üzerinde fazla bir etkisi yoktur. Bu nedenle kroma vektörünün oktav-değişmez özellikleri onu bu tarz işlemler için uygun kılar [27]. [47-53] nolu çalışmalar, akor tanıma ile ilgili, kroma yapısı üzerinden yapılan önemli çalışmalardır. Bu çalışmalarda esas hedef, sistemin öğrenme ve sınıflandırma yönlerini geliştirmektir. Örneğin [47] nolu çalışma, tam zamansal hizalamaları olmayan kopyalanmış akor dizileriyle, müzik sesi arasındaki doğrudan analogiye dikkat çekmektedir. Konuşma algılama yapısını eğitmek için kullanılan hizalanmamış kelime transkripsiyonları ile konuşma kayıtlarını, Baum-Welch prosedürü ile, her bir akorun özellikleri için her iki modeli ve eğitim verilerindeki etiket hizalamasını eşzamanlı olarak tahmin etmek için kullanmıştır. [53] nolu çalışmada, manuel olarak etiketlenmiş akor transkripsiyonlarının artan hacminden yararlanılmıştır. Böylece [53] nolu çalışma neticesinde, [52] nolu çalışmada uygulanan tonal (sinüzoidal) ve algılayıcı (geçici veya gürültülü) ayrımı yöntemine ve [46] nolu çalışmada gerçekleştirilen, enerji ve kroma hesaplaması için önceden hesaplanan yerel spektral normalizasyon hesabı yöntemine katkılar sağlanmıştır [27].

Bu alanda çığır açan çalışmalardan biri, Avery Wang tarafından 1999 yılında önerilen, yerel ses özelliği çıkarımına dayanan bir ses kopya algılama algoritması olan Shazam algoritmasıdır [3]. Bu algoritma, bir ses verisine ait spektrogram çıktısı üzerindeki kilit noktalar olarak tanımlanabilecek baskın noktaları (en yüksek enerjiye sahip noktaları) kullanarak bu veriye özel spesifik veriler üretir. Baskın noktalar Wang’a ait çalışmada çiftler halinde kullanılmıştır. Bu noktalara ait zaman ve frekans değerlerindeki farklılıklarla türetilen hash değer verileriyle ses dosyalarının benzerlik karşılaştırma işlemleri gerçekleştirilmiştir. Ses parmak izi olarak tanımlanan bu yapılar sayesinde eldeki mevcut verilerin entropisi arttırılarak daha detyalı ve kesin sonuçlara ulaşmak amaçlanmıştır. Bu yöntemle çıkarılan parmak izleri, çeşitli gürültü türlerine karşı oldukça dayanıklıdır.

Bununla birlikte, baskın noktaların konumlarına olan etkilerinden dolayı, algoritma perde veya tempo deęişikliklerine karşı çok başarılı olamamaktadır.

2.2 Ses Sinyali

Bir ses sinyalini dinlediğimizde, kulaklarımızın algıladığı şey aslında bir dizi hava basıncı deęişikliğidir. Frekans ve genlik bir ses dalgasının en önemli iki karakteristiğidir. Bunlardan frekans, ses dalgasının bir saniye içerisinde kaç defa kendisini tekrar ettiğini betimler. Birimi Hz'dir. Genlik ise, hava basıncının gücüdür. Sıfır genlik, hava basıncının normal olduğunu, pozitif genlik değeri ise hava basıncının normalden daha güçlü olduğunu ve negatif genlik değeri de hava basıncının normalden daha zayıf olduğu anlamlarına gelmektedir. Algısal bir perspektiften bakılırsa, frekans ve genlik sırasıyla perde ve ses yüksekliği ile ilişkilidir, ancak aralarındaki ilişki doğrusal değildir [54]. Dijital formda bir ses sinyali örneęi Şekil 2.4 te gösterilmiştir.



Şekil 2.4 : Bir ses sinyali örneęi.

Bir ses dalgasını işlemek için ilk adım, bir analog-dijital dönüştürücü kullanarak bu ses dalgasını dijital hale getirmektir. Aslında burada iki aşama vardır. Bunlar örnekleme ve nicemleme (kuantalama) adımlarıdır. Örnekleme, bir saniyede alınan örnek sayısı olarak

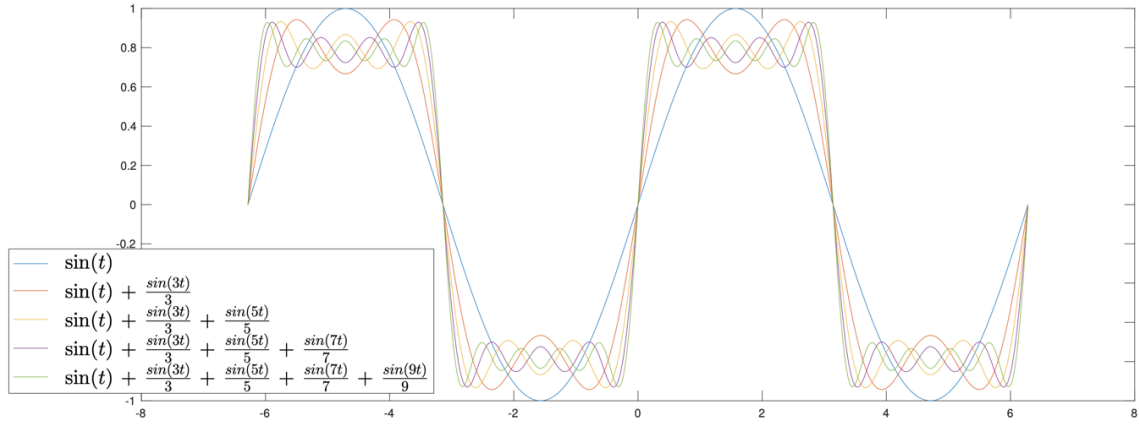
tanımlanır. Aslında uygulanan işlem belirli bir örnekleme hızıyla ses dalgasının genliğini ölçmek ve bu değerleri kaydetmektir. Nyquist-Shannon örnekleme teoremine göre, örnekleme frekansı, yakalamak istediğimiz sinyalin sahip olduğu bileşenlere ait frekanslardan, maksimum frekansın en az iki katı değerinde olmalıdır [55]. Bu durum denklem 2.1 de gösterilmiştir. Bu teorem aslında, Claude Shannon'ın bilgi teorisi üzerine yaptığı çığır açan çalışmanın önemli bir sonucudur. Bu çalışmanın sonucunda, bir analog sinyalin, ancak maksimum frekans içeriğinin iki katından daha yüksek bir frekansta örneklendiğinde tam olarak yeniden oluşturulabileceği ortaya konulmuştur.

$$F_{max} < \frac{F_s}{2} \quad (2.1)$$

Burada F_{max} , analog sinyal içerisindeki maksimumum frekansa sahip bileşenin frekans değeri ve F_s ' de örnekleme frekansıdır ve bu terim aynı zamanda Nyquist frekansı olarak da bilinir. Bu şart sağlanmadığı takdirde, elde edilen sinyal ile analog sinyal arasında bir örtüşme sağlanamaz.

2.3 Fourier Analizi

Bir ses sinyalinin zaman domeninden frekans domenine taşınmasını sağlayan matematiksel işleme dönüşüm adı verilir. Sinyal işleme alanındaki en önemli dönüşüm tekniği Fransız matematikçi Joseph Fourier tarafından 1807 yılında oluşturulan ve kendi adıyla anılan Fourier dönüşümüdür. Fourier, bu dönüşüm ile herhangi bir periyodik sinyalin, sinüs sinyallerinin toplamı biçiminde ifade edilebileceğini ispatlamıştır. Bu teknik, ses sinyali işleme alanında frekans analizi için oldukça hassas sonuçlar vermektedir. Bu sebeple en sık kullanılan yöntemlerden biridir. Bu dönüşüme ait en temel gösterimlerden biri Şekil 2.5 te verilmiştir. Bu gösterimde, periyodik bir kare dalga sinyalin sinüs bileşenleri türünden analizi adım adım gösterilmiştir. Şekilde harmonik bileşenlerin sayısının artmasıyla, toplam sinyalin giderek bir kare dalgaya yakınsadığı açıkça görülmektedir.



Şekil 2.5 : Bir kare dalganın sinüzoidal sinyaller ile ifadesi [56].

Ses sinyali işleme alanında spektrum kavramı, bir ses dalgasının frekans domenindeki tüm frekans bileşenlerinin bir temsidir. Spektrum çözünürlüğü, hangi dönüşümün kullanıldığına, örnekleme frekansına ve spektrumu hesaplamak için kaç örnek kullanıldığına bağlıdır.

Bir sinyale ait spektrum verisi, sinyale ait bileşenlere Fourier dönüşüm uygulanarak elde edilir. Bunun için en çok kullanılan yöntem Ayırık Zamanlı Fourier Dönüşümü'dür (Discrete Fourier Transform-DFT). Çalışma sırasında ayırık zamanlı sinyallere Fourier analizinin bu türü uygulanmıştır.

2.3.1 Ayırık zamanlı Fourier dönüşümü

Ayrık zamanlı Fourier dönüşümü, Fourier dönüşüm tekniğinin gerçek uygulamalarda en çok kullanılan yöntemidir. Ayırık zamanlı sinyallere uygulanan bu dönüşüm ile sinyale ait spektrum verileri elde edilir. Ayırık zamanlı Fourier dönüşümü Denklem 2.2 de gösterildiği gibi hesaplanır [56].

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-jk\frac{2\pi}{N}n} \quad (2.2)$$

$$0 \leq n < N, 0 \leq k < N \text{ ve } j \equiv \sqrt{-1}$$

Burada x_n , ses sinyaline ait giriş dizisi, X_k ise frekans çıkışıdır. N ise işlem yapılacak örnek sayısıdır.

Ters ayırık zamanlı Fourier dönüşümü (Inverse Discrete Fourier Transform-IDFT) ise denklem 2.3 de gösterildiği gibi ifade edilir:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cdot e^{jk\frac{2\pi}{N}n} \quad (2.3)$$

$$0 \leq n < N, 0 \leq k < N \text{ ve } j \equiv \sqrt{-1}$$

Bu dönüşüm tekniği, sürekli zamanlı aperiodyik sonlu enerjili sinyallerin dönüşümü için kullanılır.

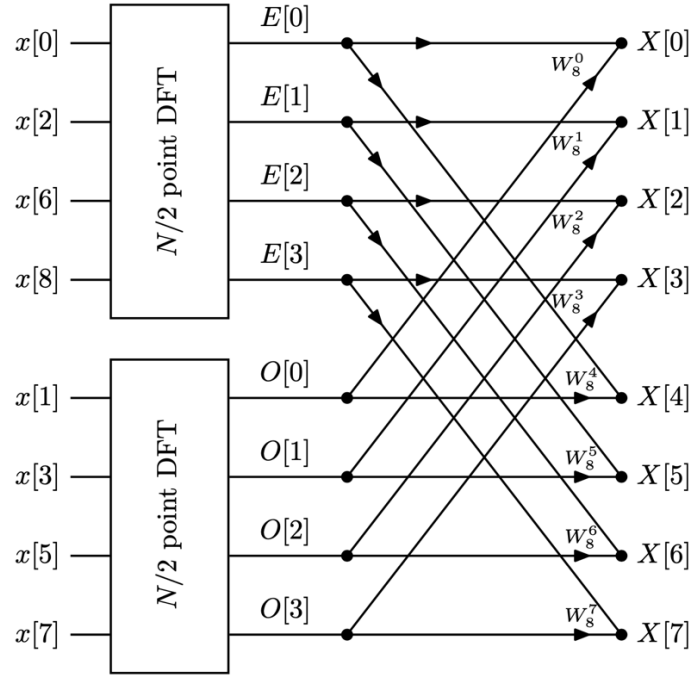
Fourier dönüşümü uygularken dikkat edilmesi gereken önemli iki noktadan birincisi şu şekilde açıklanabilir: Bir sinyalin örnekleme frekans değeri F_s büyüdükçe frekans eksenindeki örnekleme yoğunluğu artacaktır (F_s/N). Bu durumda, işleme dahil edilecek örnek sayısı artmış ve dönüşüm daha fazla nokta dikkate alınarak sürdürüleceği için, frekans ekseninde daha hassas sonuçlar alınacaktır. Yani frekans çözünürlüğü artacaktır.

İkinci önemli nokta ise şu şekildedir: Bu dönüşüm sinyalin zamanda değişmez bir yapıda olduğunu kabul etmektedir. Örneğin, periyodik sinyallerin toplamından oluşan bir sinyale bu dönüşüm uygulanırsa, bu yöntem bu sinyal için doğru sonuçlar verecektir. Diğer taraftan, işlenecek sinyal, yine benzer şekilde periyodik sinyallerin toplamı şeklinde devam ederken, bir noktada başka bir periyodik veya periyodik olmayan bir sinyal bu toplam sinyale eklendiği zaman, artık bu sinyal durağanlık özelliğini kaybetmiş olacaktır. Fourier dönüşümüne göre dönüşümü yapılan bu sinyal, artık eklenen sinyalin eklendiği yere kadar durağandır. Bu noktada, artık frekans çözünürlüğü değişmiştir ve bu istenmeyen bir durumdur. Çünkü, kullanılan dönüşüm tekniğinin gerektirdiği şart sağlanmadığı için artık hatalı sonuçlar oluşacaktır [57]. Bu sebeple sinyaller üzerinde işlem yapılırken sinyaller parçalar halinde işlenerek bu durumun önüne geçilmeye çalışılır. Sinyalin durağan olduğu aralıklar belirlenebilirse, hatalı sonuçlar elde edilmez. Bu noktada bu aralığın tespiti yeni bir zorluk olarak karşımıza çıkmaktadır. Çünkü ses sinyalleri, yapıları gereği çok geniş bir spektrum yelpazesinde meydana gelirler ve aperiodyik sinyallerin durağan oldukları bir aralığı belirlemek pratikte çok mümkün değildir. Bunun yerine belirli aralıklar dahilinde bu işlemler yürütülerek en iyi sonuçlara ulaşılmaya çalışılır. Bu durum bu çalışmada ele alınan konulardan birisidir. Metodoloji ve uygulama bölümlerinde bu konu pencere boyu kavramı ile detaylandırılmıştır.

2.3.2 Hızlı Fourier dönüşümü

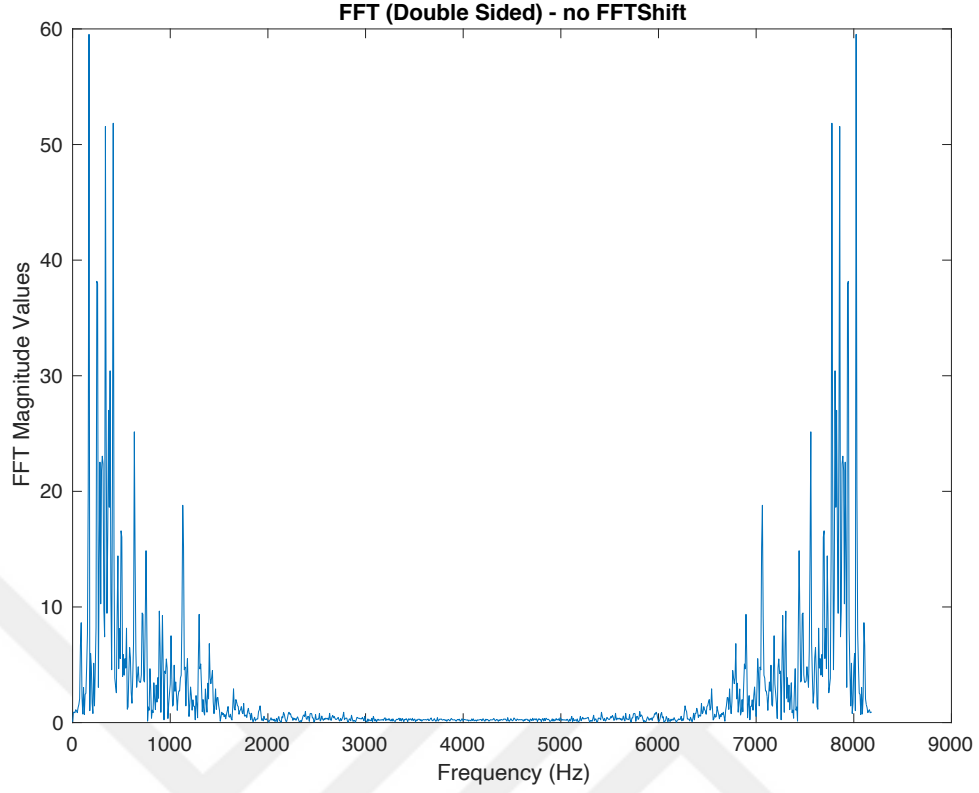
Ayrık zamanlı Fourier dönüşümün hesaplanması için çeşitli algoritmalar geliştirilmiştir. Bu algoritmalar içerisinde, hız optimizasyonu bakımından en başarılı

algoritmalarından biri, literatürde Radix-2 decimation-in-time olarak geçen, Cooley ve Tukey tarafından oluşturulmuş FFT (Fast Fourier Transform – Hızlı Fourier Dönüşümü) algoritmasıdır [58]. Bu algoritma, uzunluğu N olan bir DFT'yi alır ve yinelemeli olarak iki eşit boyutlu (başlangıçta $N/2$ uzunluğunda) serpiştirilmiş DFT'ye böler. 8 FFT uzunluğunda yürütülen bir FFT örneği Şekil 2.6 da ki gibidir.



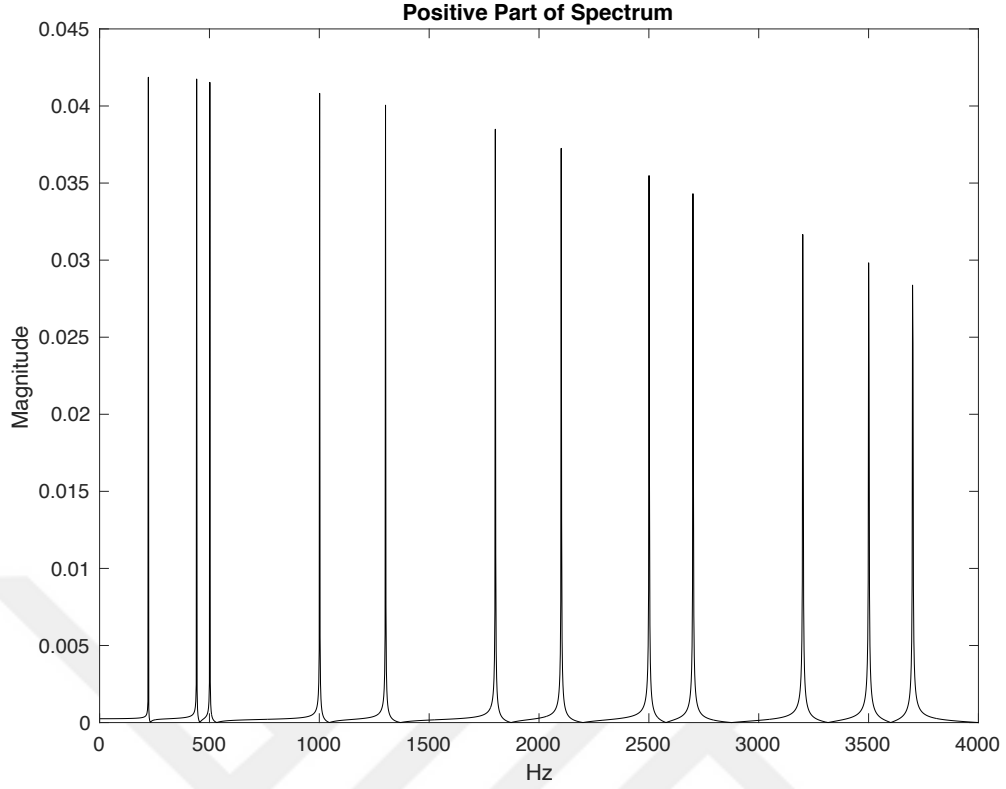
Şekil 2.6 : Cooley ve Tukey tarafından oluşturulmuş tek aşamalı 8 birim uzunlukta Radix-2 FFT'nin devre şeması [56].

Matlab'da bu işlemi gerçekleştirebilmek amacıyla, sinyal işleme kütüphanesinde bulunan `fft()` fonksiyonu kullanılmaktadır. Çalışma boyunca bu fonksiyondan yararlanılmıştır. Bu fonksiyonun kullanımına ait bir örnek çıktı Şekil 2.7 de verilmiştir. Bu işlem için bir çoklu ortam ses dosyasına ait veriden 8192 örnek uzunluğunda bir parça kullanılmıştır. Şekilden de görüleceği üzere dönüşüme ait çıktı simetrik bir yapıdadır. Bunun sebebi, Fourier dönüşümünün gerçek sinyallere uygulanması sırasında, pozitif ve negatif frekans spektrumu bileşenlerinin eşit değerlerde elde edilmesidir. Yani simetrik yapıda bir sonuç elde edilir. İşlemler gerçekleştirilirken bu durum göz önünde bulundurulup, çıkan sonucun yarısı üzerinden değerlendirmeler sürdürülür.



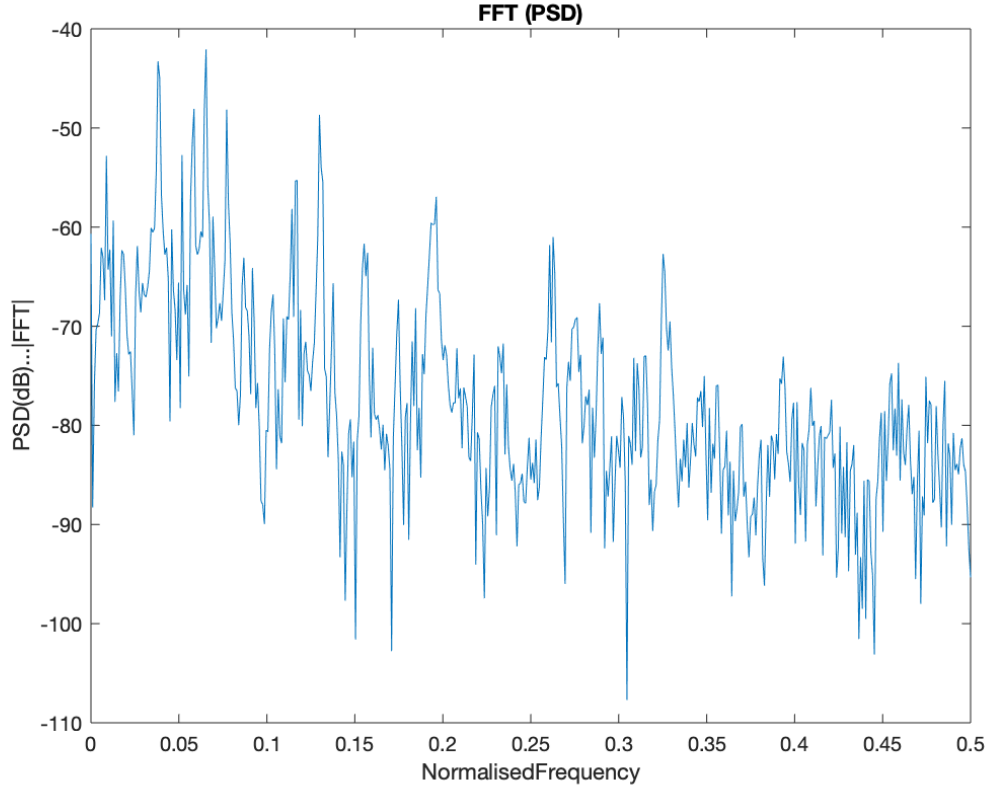
Şekil 2.7 : 8192 örnekten oluşan bir ayırık zamanlı sinyale ait FFT grafiği.

Bu fonksiyonun kullanımına ait bir başka örnek Şekil 2.8 de verilmiştir. Bu defa dönüşüm işlemine tâbi tutulan sinyal periyodik sinüzoidal sinyallerin toplamından oluşan periyodik bir sinyaldir. Yani dönüşümün en net sonuçları verebileceği türden bir sinyal işleme konulmuştur. Şekilden görüleceği üzere bileşenlere ait frekans değerleri spektrum ekseninde oldukça net bir şekilde görülmektedir.



Şekil 2.8 : 220, 440, 500, 1000, 1300, 1800, 2100, 2500, 2700, 3200, 3500, 3700 Hz'lik sinüzoidal sinyallerin toplamından oluşan bir sinyale ait spektrum grafiği.

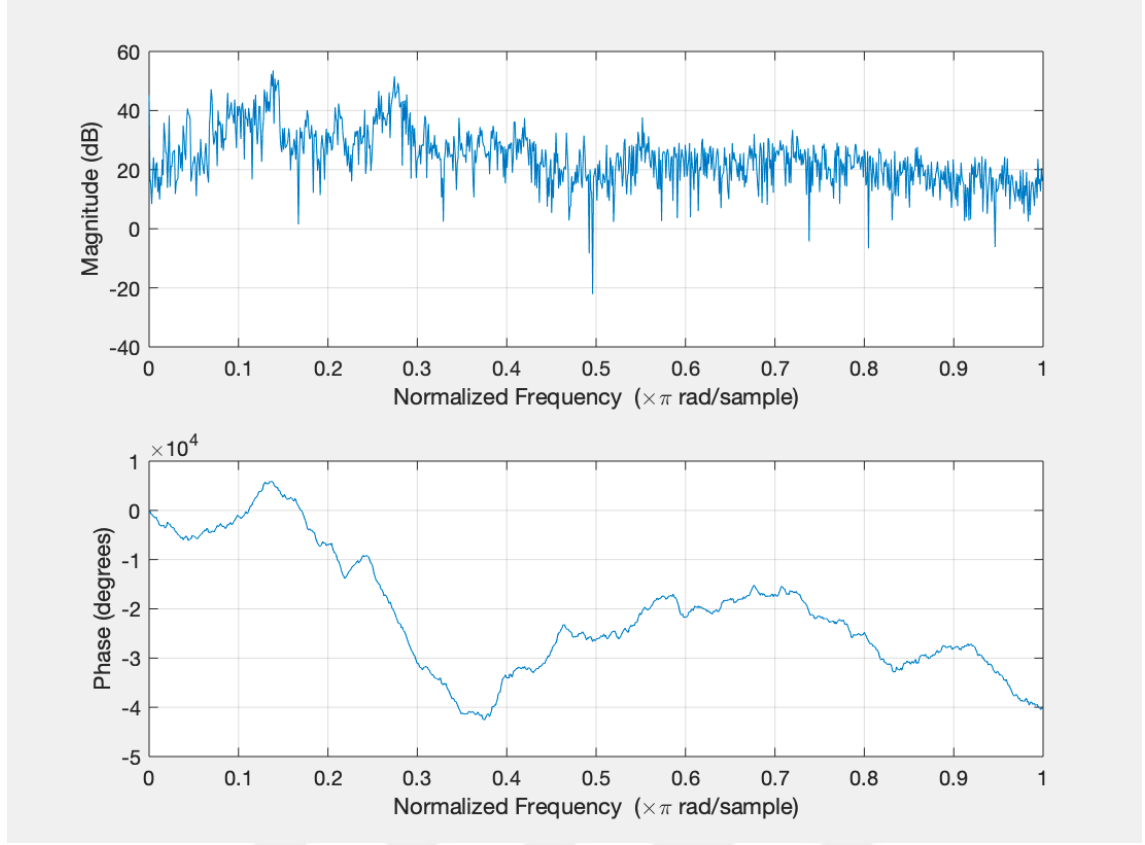
Bir sinyale ait frekans bileşenleri hakkında fikir sahibi olabileceğimiz bir başka teknik ise sinyale ait spektral güç yoğunluğu (Power Spectral Density-PSD) değerlerinin analizidir. Bu işleme ait bir örnek Şekil 2.9 da verilmiştir. Üstte bahsedildiği üzere simetrik olarak elde edilen verilere ait çıktının yarısı, normalize edilmiş bir frekans skalasında çizdirilmiştir.



Şekil 2.9 : 1024 örnek uzunluğundaki bir ses sinyaline ait FFT dönüşümüne ait PSD çıktısı.

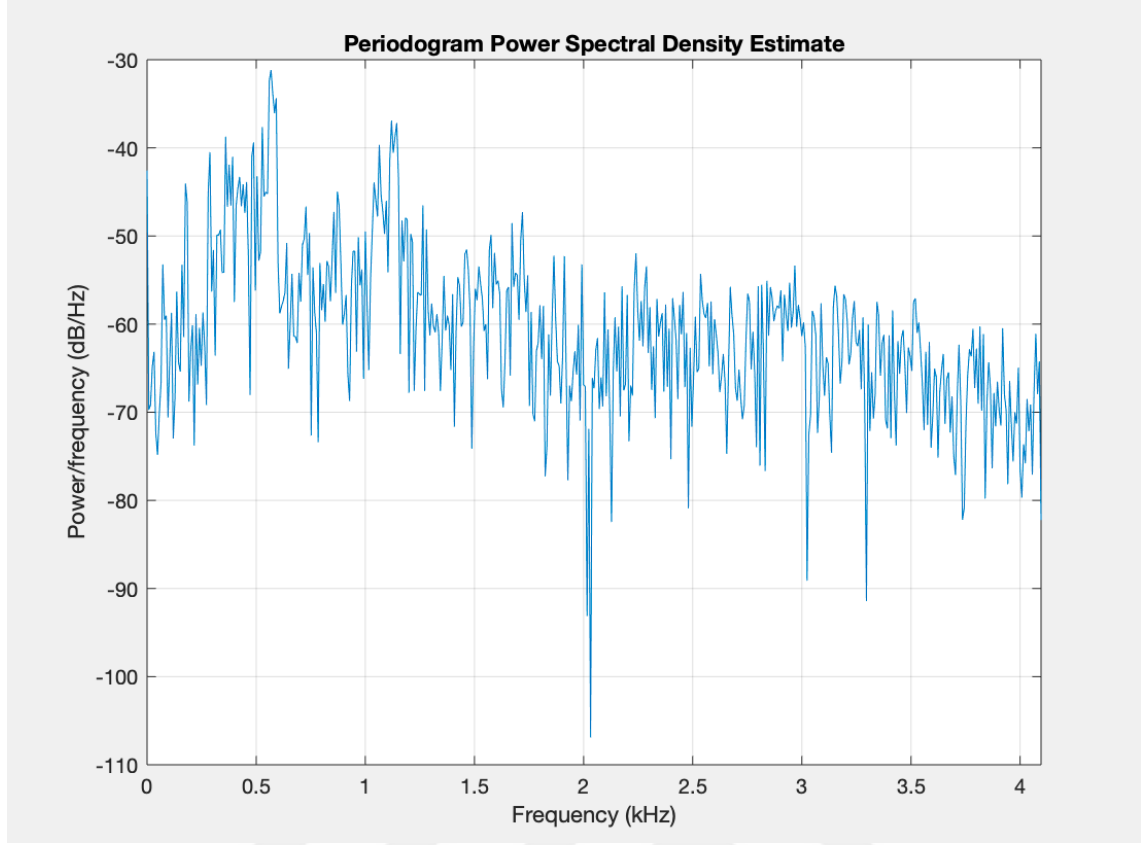
Spektral güç yoğunluğu, sinyale ait frekans bileşenlerinin genlik değerlerinin karesini temsil eden bir değerdir. Bunun için Matlab üç adet hazır fonksiyon sunmaktadır. Bunlar; `freqz`, `periodogram` ve `pwelch` fonksiyonlarıdır.

Bu fonksiyonlardan `freqz` bir rastgele IIR (Infinite Impulse Response-Sonsuz Dürtü Yanıtı) filtrenin frekans analizi için kullanılabildiği gibi, gerçek bir sinyalin tek taraflı frekans bilgisini (FFT verisi) elde etmek için de kullanılabilir [56]. Bu fonksiyon yardımıyla bir ses sinyaline ait hesaplanmış FFT-genlik ve FFT-faz değerlerine ait veri Şekil 2.10 da verilmiştir.



Şekil 2.10 : Matlab `freqz()` fonksiyonu yardımıyla bir sinyale ait FFT genlik ve FFT faz çıktısı [56].

Bu fonksiyonlardan bir diğeri olan `periodogram()` ise daha çok bir sinyale ait spektral güç yoğunluğunu hızlı bir şekilde hesaplamak için kullanılır. Bu fonksiyon da `freqz()` fonksiyonunda olduğu gibi bu işlemi yaparken FFT genlik değerlerinden bu hesaplamayı gerçekleştirir ve benzer sonuçlar elde edilir. Bu fonksiyon, `freqz()` fonksiyonundan farklı olarak, gerçek sinyallerin yanında kompleks sinyallere ait çift taraflı frekans analizini de mümkün kılar. Bu fonksiyonun kullanımına ait bir çıktı Şekil 2.11 de verilmiştir.



Şekil 2.11 : Bir sinyale ait `periodogram()` fonksiyonu yardımıyla hesaplanmış FFT genlik verisi [56].

Daha önce de bahsedildiği üzere, ayrık zamanlı Fourier dönüşüm tekniği, üzerinde işlem yapacağı sinyali eksi sonsuz ile artı sonsuz zaman aralıklarında, sinyalin belirli bir periyot ile kendisini tekrar ettiğini varsayarak işlemlerini gerçekleştirmektedir. Dijital sinyal işlemede uygulanan en temel yöntemlerden birisi de sinyalin bir bütün olarak değil, parçalar halinde işlenmesi tekniğidir. Bu parçalar hızlı Fourier dönüşümü işlemine tâbi tutulduğu zaman, dönüşüm, yukarıda belirtilen varsayımla, işlenen sinyalin periyodik ve eksi ile artı sonsuz zaman aralığında kendisini tekrar ettiğini varsayarak işlemlerini gerçekleştirecek ve sonuçlarını bu varsayıma göre oluşturacaktır. Ancak gerçek hayattaki uygulamalarda bu şartları sağlayan sinyaller çok nadir bulunur. Dolayısıyla bu sinyallere ait hesaplamalardan hatalı sonuçlar elde edilecektir. İşte bu hatalı varsayım ile dönüşüm, alınan her bir sinyal parçasını sonsuz zaman aralığında, işleme alınan uzunluğu kadar periyotta bir periyodik sinyal olduğunu varsayarak işlemlerini gerçekleştirir. İşlemler sonucunda elde edilen spektrum sinyallerinin yan yana birleştirilmesi esnasında, birleşim (eklem) noktalarında bir örtüşme problemi meydana gelmekte ve bu sebeple hatalı sonuçlar ortaya çıkmaktadır. Literatürde bu olay spektral sızıntı olarak geçmektedir. Yani sinyale ait frekans bileşenlerinin eklem noktalarında meydana gelen uyuşmazlık neticesinde frekans

özünürlüğünde kayıplar meydana gelmektedir. Bu sızıntı tamamen engellenememekle birlikte, çeşitli pencere fonksiyonlarının dikkatli bir şekilde uygulanmasıyla, önemli ölçüde azaltılabilir [56].

2.4 Pencere Fonksiyonları

Pencere fonksiyonları, zaman-frekans domeninde sinyal üzerinde işlem yapılırken, özellikle sinyale ait spektrogram oluşturulurken, dönüşümün kalitesinin artırılması için sıklıkla kullanılan yapılardır. Frekans bilgisine yönelik yapılan hesaplamalarda, sinyale ait kısa zamanlı örneklerin Fourier dönüşümleri pencere fonksiyonları yardımı ile spektrogram verilerine dönüştürülerek veriler üzerinden işlemler gerçekleştirilir.

Pencere fonksiyonları, zaman domenindeki verilere uygulanan sınırlı ve değişmez yapıdaki ağırlık fonksiyonları olarak tanımlanabilir. Sonlu dürtü yanıtı (FIR-Finite Impulse Response) filtrelerinin tasarımı, görüntü işleme, konuşma tanıma, spektrum analiz gibi sinyal işleme alanının en temel uygulamalarında oldukça önemli bir rol oynarlar.

Pencere fonksiyonlarının sinyal işleme alanındaki en temel iki uygulaması: FFT temelli veri analizi ve IIR (Infinite Impulse Response-Sonsuz Dürtü Yanıtı) filtrelerden FIR filtre tasarımıdır. FFT ile veri analizindeki kullanım amacı, bu çalışmada da uygulanan spektral sızıntının azaltılması amacına yöneliktir. FIR filtre tasarımı ise Gibbs salınımlarını azaltmaya yönelik kullanılırlar [60].

Üzerinde işlem yapılacak bir sinyalden herhangi bir parçasının seçim işleminde yine pencere fonksiyonları kullanılır. Yapılan pencere seçimi, üzerinde işlem yapılacak sinyal parçasına ait farklı sinyal parametrelerinin hangi kalitede elde edileceğini etkiler [61, 62]. Pencere seçimi yapılırken, amaca uygun pencere modelinin hangisi olacağı, pencerenin uzunluğunun ne kadar olacağı, sinyal üzerinde işlem yapılırken pencerenin hangi konumda uygulanacağı, üst üste gelme oranının ne olacağı gibi konuların önemle dikkate alınması gerekir.

Bir pencere fonksiyonunun frekans domenindeki istenen karakteristik yapısı, ana lobun dar bir yapıda ve yan lobların ise dar ve hızlı sönümlenir bir yapıda olması şeklindedir. Ancak bu iki istek birbiriyle çelişkilidir. Çünkü belirli uzunluktaki bir pencere fonksiyonu ne kadar dar bir ana loba sahip olursa, istenen hızlı sönümlenme özelliğinden bir o kadar uzaklaşmış olur. Tam tersi şekilde, yan lobların hızlı sönümlenmesi durumunda, ana lob istenilen aksine daha geniş yapıda olmaktadır.

Matlab, bu konuda sinyal işleme kütüphanesinde kullanıma hazır bir takım pencere fonksiyonları barındırmaktadır. Bunlardan bazıları Çizelge 2.1 de verilmiştir.

Çizelge 2.1 : Bazı yaygın kullanılan Matlab pencere fonksiyonları.

Komut Adı	Komuta ait pencere fonksiyonu
barthannwin	Modified Bartlett-Hann window
bartlett	Bartlett window
blackman	Blackman window
blackmanharris	Minimum four-term Blackman-Harris window
bohmanwin	Bohman window
chebwin	Chebyshev window
enbw	Equivalent noise bandwidth
flattopwin	Flat top weighted window
gausswin	Gaussian window
hamming	Hamming window
hann	Hann (Hanning) window
kaiser	Kaiser window
nuttallwin	Nuttall-defined min 4-term Blackman-Harris window
parzenwin	Parzen (de la Vallée Poussin) window
rectwin	Rectangular window
taylorwin	Taylor window
triang	Triangular window
tukeywin	Tukey (tapered cosine) window

Bu çalışmada kullanılan bazı pencere fonksiyonlarının detaylı özellikleri aşağıda sıralanmıştır.

2.4.1 Flat-top penceresi ve özellikleri

Flat-top penceresi adını ana lob yapısındaki düzlükten dolayı almıştır. Üçüncü ve beşinci dereceden Flat-top penceresi eşitlikleri Denklem 2.4 te verilmiştir [63].

$$f[n] = a_0 + a_1 \cdot \cos\left(\frac{2\pi n}{N}\right) + a_2 \cdot \cos\left(\frac{4\pi n}{N}\right) + a_3 + a_4 \cdot \cos\left(\frac{8\pi n}{N}\right) \quad (2.4)$$

$$0 \leq |n| \leq \frac{N}{2}$$

Üçüncü derece Flat-top penceresi için katsayılar:

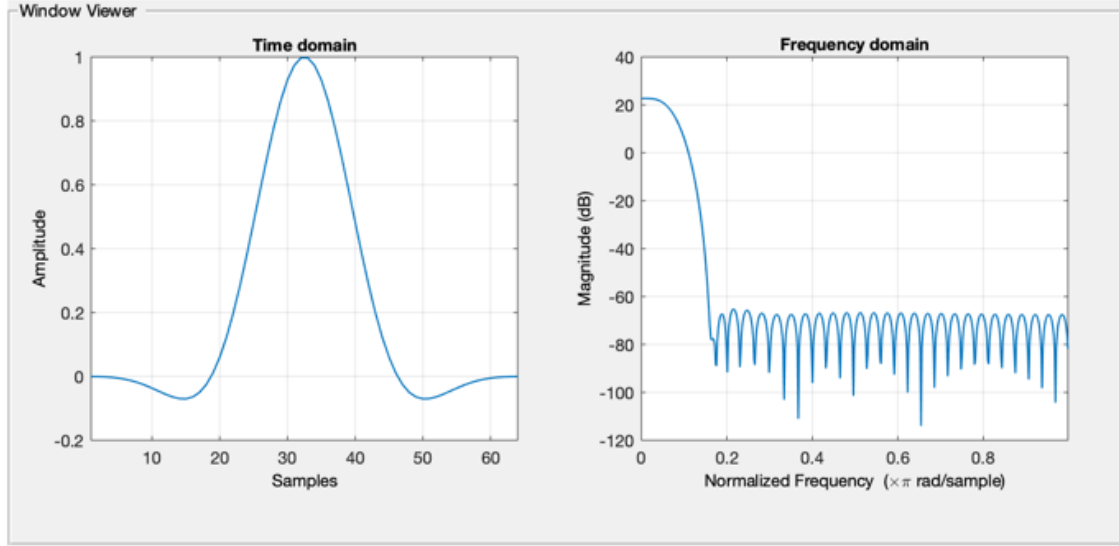
$a_0 = 0.2811, a_1 = 0.5209, a_2 = 0.198$ şeklindedir.

Beşinci derece Flat-top penceresi için katsayılar:

$a_0 = 0.21557895, a_1 = 0.41663158, a_2 = 0.277263158, a_3 = 0.083578947,$

$a_4 = 0.006947368$ şeklindedir.

Zaman ve frekans domeninde 64 örnek uzunlukta bir Flattop penceresi Şekil 2.12 de gösterilmiştir.



Şekil 2.12 : 64 örnek uzunlukta Flattop pencere fonksiyonunun zaman ve frekans domeninde gösterimi.

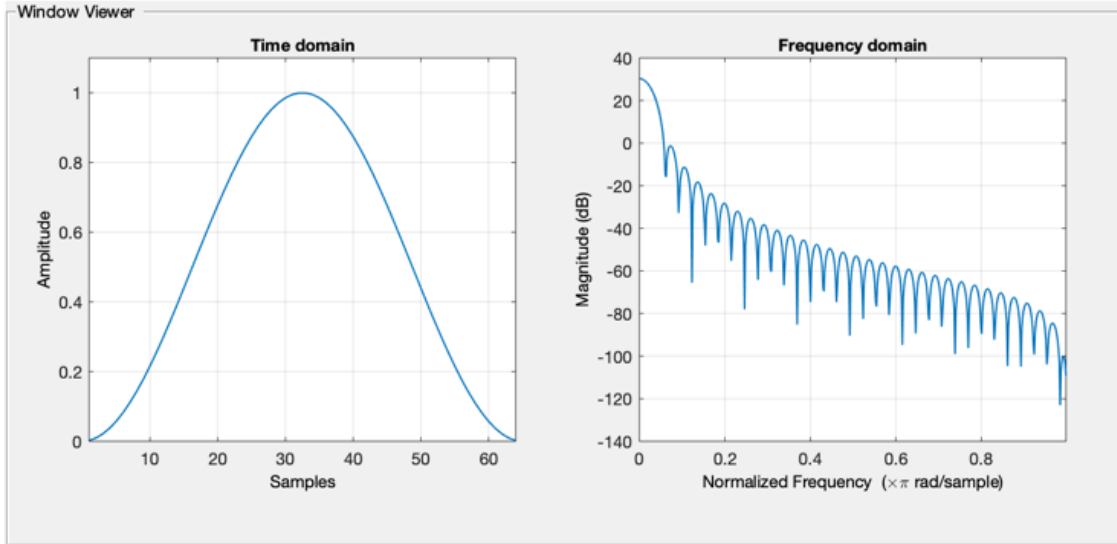
2.4.2 Hanning penceresi ve özellikleri

Hanning penceresi aslında $\cos^2 x$ penceresinin özel bir halidir. “a” değişkeninin değeri 2 alınarak oluşturulmuştur ve ifadesi Denklem 2.5 te ki gibidir [64].

$$f[n] = 0.5 + 0.5 \cdot \cos\left(\frac{2\pi n}{N}\right) \quad , \quad 0 \leq |n| \leq \frac{N}{2} \quad (2.5)$$

Hanning penceresinde ilk yan lob dikdörtgen pencereye nazaran daha zayıf ve daha hızlı sönümlenme hızına sahiptir. Bu, 60 dB bant genişliğinin çok daha dar olduğu ve çok daha iyi seçicilik sağladığı anlamına gelir. “Picket Fence Effect” olarak bilinen ve geçiş bandındaki dalgalanma tarafından verilen maksimum genlik hatası, bu pencere için yalnızca 1.4 dB’dir [65].

Zaman ve frekans domenlerinde 64 örnek uzunlukta bir Hanning penceresi Şekil 2.13 de verilmiştir.



Şekil 2.13 : 64 örnek uzunlukta Hanning pencere fonksiyonunun zaman ve frekans domeninde gösterimi.

2.4.3 Dikdörtgen pencere ve özellikleri

Literatürde düz ya da boxcar ağırlıklandırma olarak geçen dikdörtgen pencere aslında sonlu zaman aralığında herhangi bir ağırlıklandırma sağlamaz. Deterministik veya harmonik sinyallerin analizinde bu pencere zayıf bir filtre özelliği gösterir. Çünkü, bu pencere 60 dB bant genişliği ile çok zayıf bir seçicilik sağlayacaktır. Göstermiş olduğu bir diğer dezavantaj ise geçiş bandında gösterdiği nispeten büyük bir (3.9 dB) dalgalanma karakteristiği sunmasıdır [65].

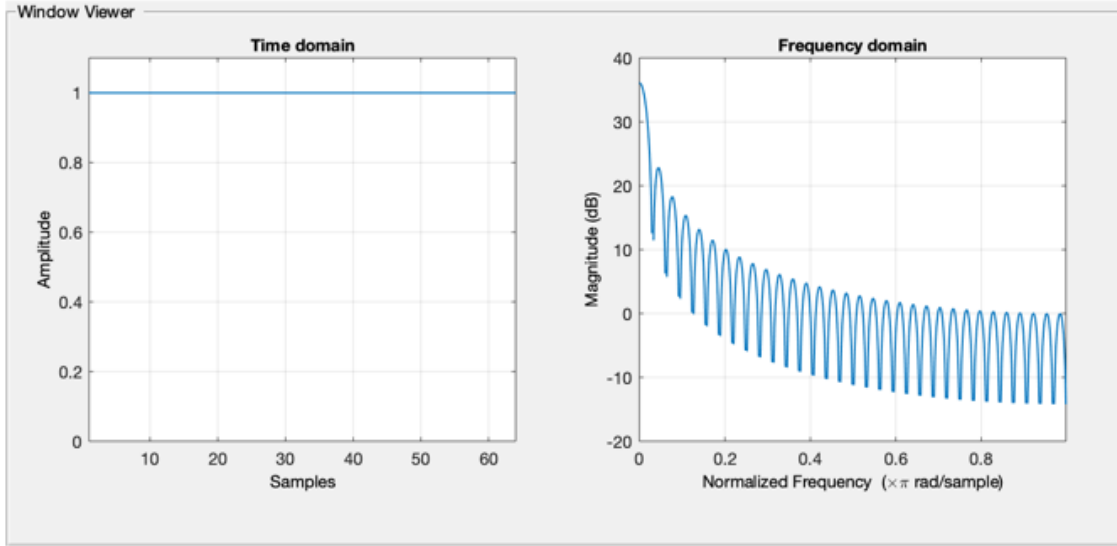
Dikdörtgen pencerenin eşitliği denklem 2.6 da ki gibidir:

$$f(x) = \begin{cases} 1, & |t| < \tau \\ 0, & elsewhere \end{cases} \quad (2.6)$$

Bu pencerenin sürekli zamanlı Fourier dönüşümü ise sinc fonksiyonuna eşittir [66].

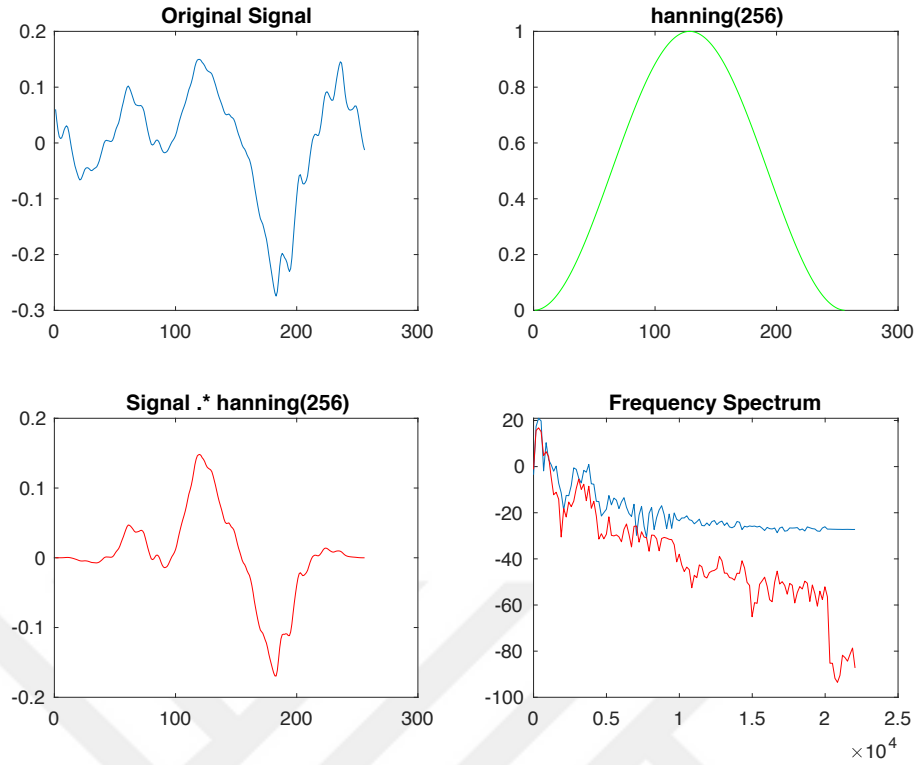
$$F(j\Omega) = \frac{2\tau \sin(\Omega\tau)}{\Omega\tau} \quad (2.7)$$

Zaman ve frekans domenlerinde 64 örnek uzunlukta bir dikdörtgen pencere Şekil 2.14 te verilmiştir.



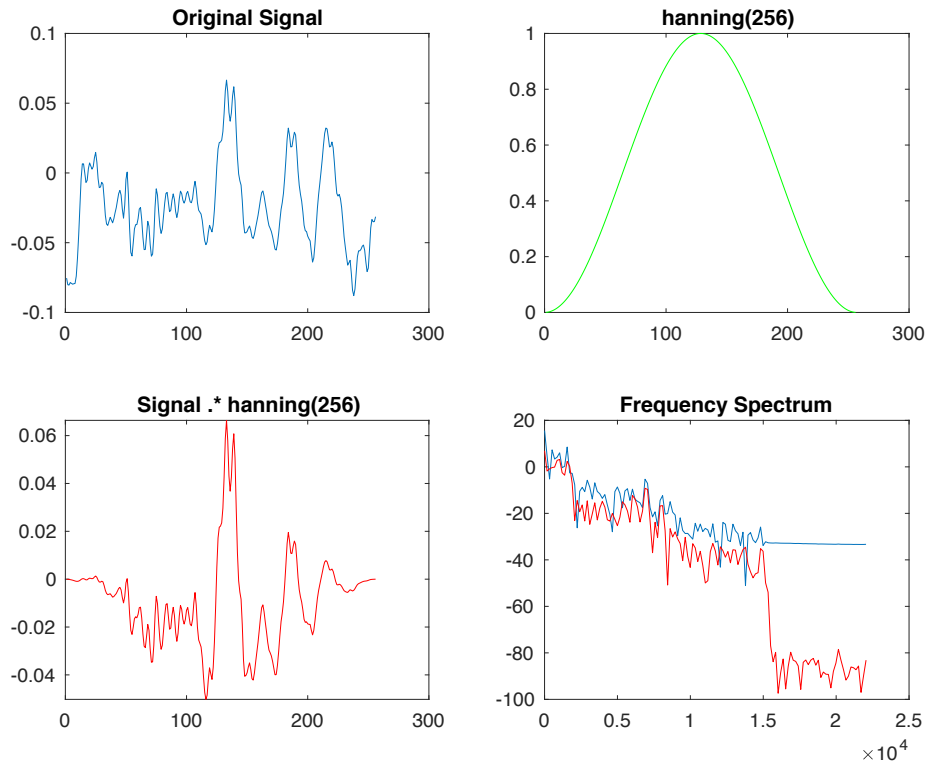
Şekil 2.14 : 64 örnek uzunlukta dikdörtgen pencerenin zaman ve frekans domeninde gösterimi.

Aşağıdaki şekillerde, pencereleme işleminin spektral etkisi basitçe gösterilmiştir. Sinyale ait orijinal veri ile elde edilen spektrum grafiği Şekil 2.15 ve Şekil 2.16 da mavi renk ile çizdirilmiştir. Pencereleme işlemine tâbi tutulan sinyal verisi ise aynı şekillerde kırmızı renk ile çizdirilmiştir. Bu verilere dikkat edilirse, pencereleme işlemine tâbi tutulmayan sinyale uygulanan FFT işlemi neticesinde elde edilen frekans çözünürlüğü, pencereleme işlemine tâbi tutulan sinyale göre daha düşük çıkmaktadır. Bunun sebebi, spektral sızıntının pencerelemenin olmadığı durumda çok daha yüksek olmasıdır. Frekans eksenindeki değerler arasında komşu değerlerine doğru erime şeklinde nitelendirebileceğimiz bir sapma yaşamıştır. Bu etki sadece sinyale ait frekans ekseninde görülmez. Aynı zamanda frekans değerlerine ait genlik değerlerinde de bir düşüşe, dolayısıyla bir hataya yol açmaktadır. Bu durum sinyal üzerinde yapılan hesaplamalarda yanlış sonuçlar oluşturacaktır. Bu sebeple uygun bir pencere fonksiyonunun seçilip kullanılması büyük önem arz etmektedir.



Şekil 2.15 : Pencereleme işleminin spektruma etkisi [56].

Bu etkinin, bir başka sinyale uygulanmış örneği şekil 2.16 da gösterilmiştir.



Şekil 2.16 : Pencereleme işleminin spektruma etkisi [56].

2.5 Spektrogram

Spektrogram, bir sinyaldeki frekans spektrumunun zamanla deęişiminin görsel bir temsidir. Ses sinyallerinin üç boyutlu bir ifade türü olarak tanımlayabileceğimiz bu yapı sinyallere ait üç boyutlu bir temsidir. Bu boyutlar: zaman, frekans ve frekans bileşenine ait genlik deęerine ait boyutlardır. Bu boyutlardan zaman boyutu x-ekseninde, frekans boyutu y-eksenindedir. Frekans genlik deęeri boyutu ise renk skalasında, renk tonlaması ile ifade edilir. Spektrogram elde etmek için kullanılan modern yöntem, dijital verideki örneklere FFT uygulanarak sinyal içerisindeki frekans spektrumunun deęer aralığının hesaplanmasına dayanmaktadır.

Fourier dönüşümü, tek başına sinyaller hakkında frekans karakteristiklerini başarılı bir şekilde oluştursa da bu frekansların hangi zamanlarda meydana geldikleri bilgisini sağlayamaz. Bu eksikliği gidermek için Gabor dönüşümü olarak bilinen STFT (Short Time Fourier Transform-Kısa süreli Fourier dönüşümü) teknięi geliştirilmiştir. Bu teknik ile kayan bir pencereye uygulanan FFT ile sinyale ait hem spektrum bilgisi hem de bu spektruma ait zaman verisi üretilir [67, 68].

Daha basit terimlerle spektrogram, yorumlanabilir bir kartezyen ızgarada en yaygın olarak ses sinyallerini temsil etmek için kullanılan bir zaman-frekans grafięidir [69]. Başka bir tanımda, bir spektrogram, bir sinyale ait kısa-süreli Fourier dönüşüm işleminin genlik deęerinin logaritmik skalada çizilmiş hali olarak belirtilmiştir. STFT verisi, sinyalin oldukça küçük parçalar (milisaniye mertebesinde) halinde FFT işlemine tâbi tutulmasıyla elde edilen dizi halindeki verilerdir. Sinyale ait işlem yapılacak bu küçük veri parçaları seçilirken, sinyalin bir önceki seçilen kısmının belli bir oranının yeniden seçilmesine dikkat edilir [70-72]. Bu işlem literatürde ‘overlapping’ olarak geçen üst üste binme oranıdır. Bu oran, optimal bir şekilde seçilmelidir. Yapılan denemelerde, bu oranın %50 seçilmesi, işlem miktarı verimi açısından optimal sonuçlar vermiştir. Bu deęerin altında seçim yapıldığı zaman, bu deęerden uzaklaştıkça sinyale ait nitelikli noktalarda azalma meydana gelmektedir. Tam tersi durumda, bu oranın üzerinde seçimler yapıldığı zaman, işlemler için gerekli olmayan noktalar işleme dahil olarak, çoęu zaman algoritmanın iş yükünü arttırmaktadır.

Bir spektrogram oluşturma teknik olarak şu şekilde gerçekleştirilmektedir. İlk olarak ses sinyali eşit parçalara ayrılır. Her bir parça spektrogramdaki bir sütuna karşılık gelecektir ya da daha net tanımlayacak olursak, belirli bir zaman anındaki, frekansa karşılık genlik

değerlerini oluşturan veri dizileri olacaktır. Bu parçalar, sinyale STFT uygulanarak elde edilen değerlerin karelerinin hesaplanmasıyla oluşturulurlar ve bu hesaplamaların sonucu yan yana dizilerek, üç boyutlu bir yapı oluşturulur [70-72, 74]. Spektrogram hesaplamasına yönelik oluşturulan eşitlik denklem 2.8 de verilmiştir [56].

$$\text{Spectrogram}(t, \omega) = |\text{STFT}(t, \omega)|^2$$

$$X_m(k) = \sum_{n=0}^{N-1} x[n + mR] \omega[n] e^{-jk\omega_0 n} \quad (2.8)$$

$$\omega_0 = \frac{2\pi}{N}$$

Burada:

$X_m(k)$: $x[n]$ giriş sinyalinin STFT değeri

m : Pencere kaydırma kontrol parametresi

R : Örnek bazında atlama miktarı

$x[n + mR]$: $n + mR$ indeksli kaydırılmış giriş sinyali

$\omega[n]$: M uzunluktaki pencere fonksiyonu, örn. Hamming

$x[n]$: n indeksli giriş sinyali

k : STFT çıkışının frekans indeksi: $k = 0, 1, \dots, N-1$

N : DFT' nin uzunluğu

Denklem 2.8 de örnek bazında atlama miktarı olarak geçen R değerinin, DFT'nin uzunluk değeri N 'den daha düşük olduğu varsayılır. İleri ayrık STFT böylelikle, m indeksi zaman eksenini ve k indeksi de frekans eksenini temsil etmek üzere, giriş sinyalinin zaman-frekans analizine ait bir matris olarak sonuçlanır.

Ters STFT (ISTFT) ise şu şekilde ifade edilir [56]:

$$y[n] = \sum_{m=-\infty}^{\infty} y_m[n - mR] \omega[n - mR]$$

$$y_m[n - mR] = \frac{1}{N} \sum_{k=0}^{N-1} Y_m[k] e^{-jk\omega_0 n} \quad (2.9)$$

$$\omega_0 = \frac{2\pi}{N}$$

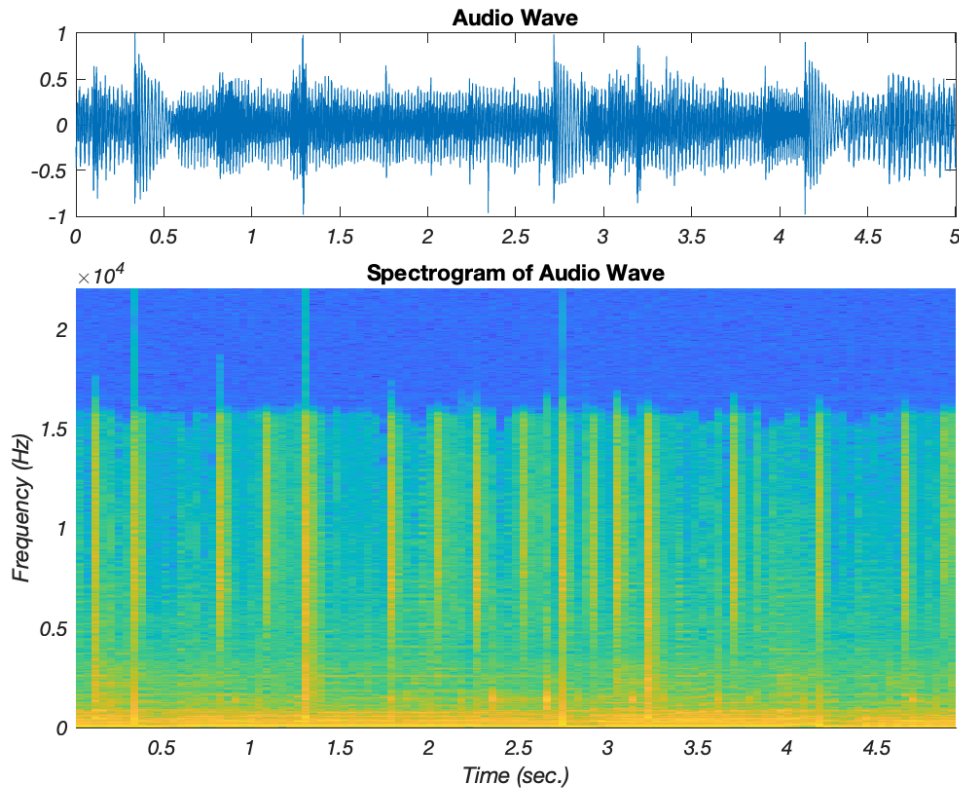
Y_m : Ters STFT işlemi için giriş sinyali

y_m : Y_m 'nin ters DFT'si

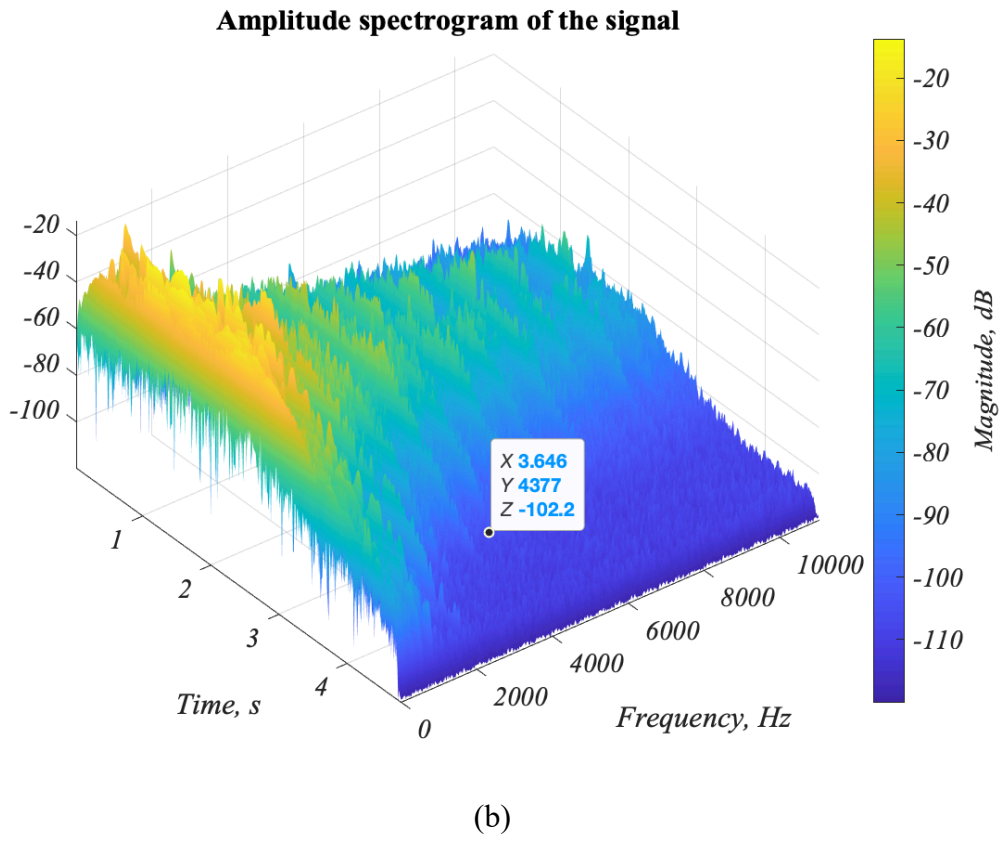
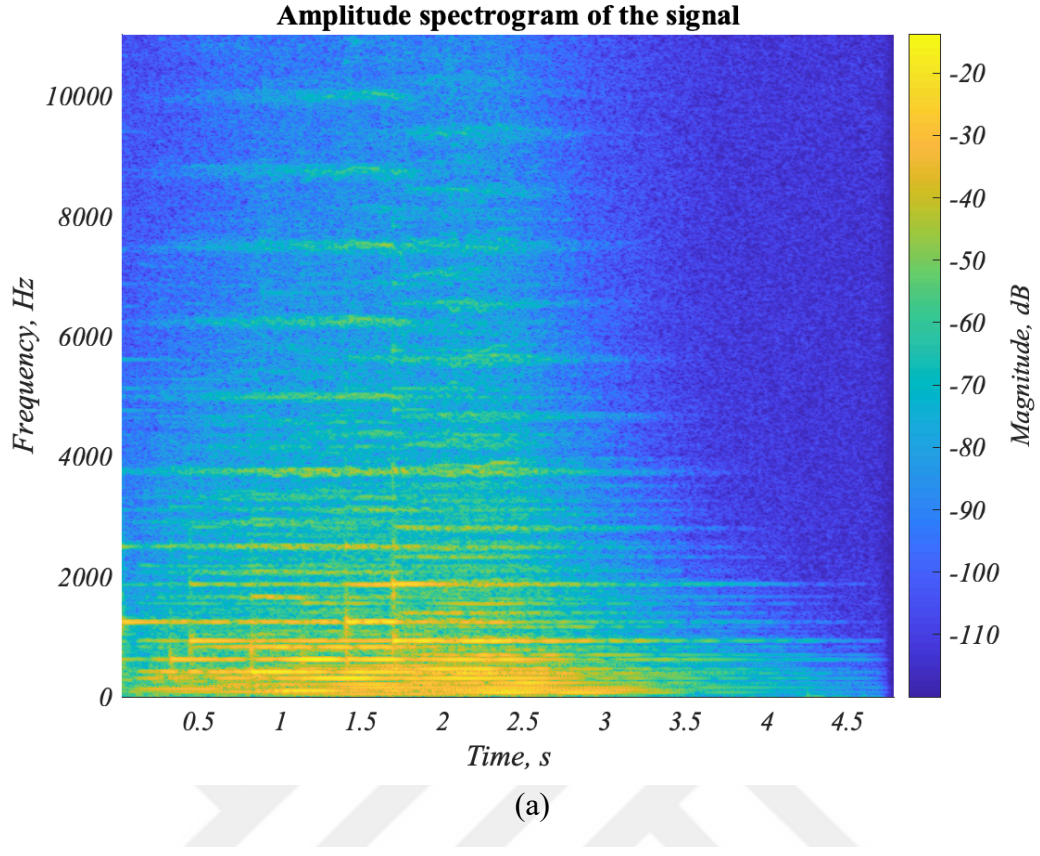
y : x 'in ters STFT'si

Matlab'da bir ses sinyaline ait spektrogramı elde edebilmek için `spectrogram()` fonksiyonu kullanılabilir. Bu fonksiyon çalıştırıldığında, giriş sinyaline ait spektrogram zaman ve frekans eksenlerinde çizdirilir ve STFT'ye ait katsayıları içeren matris elde edilebilir.

Aşağıdaki spektrogram örneği, `spectrogram()` fonksiyonu yardımıyla oluşturulmuştur. x-ekseni zaman eksen, y-ekseni frekans eksen olarak karşımıza çıkmaktadır. Zaman eksen soldan sağa doğru lineer bir şekilde artarken, frekans eksen ise aşağıdan yukarıya doğru benzer olarak lineer artmaktadır. Frekansa ait genlik değeri grafik üzerinde renklendirme uygulanarak ifade edilmiştir. Burada mavi renk tonlarına yakın renk bölgeleri, sarı renk tonuna yakın bölgelere göre genlik değeri bakımından daha düşüktür. Bir ses sinyaline ait spektrogram örneği Şekil 2.17 de verilmiştir. Benzer olarak genlik değerlerinin incelenebilmesi amacıyla Şekil 2.18 de sunulan grafik incelenebilir.



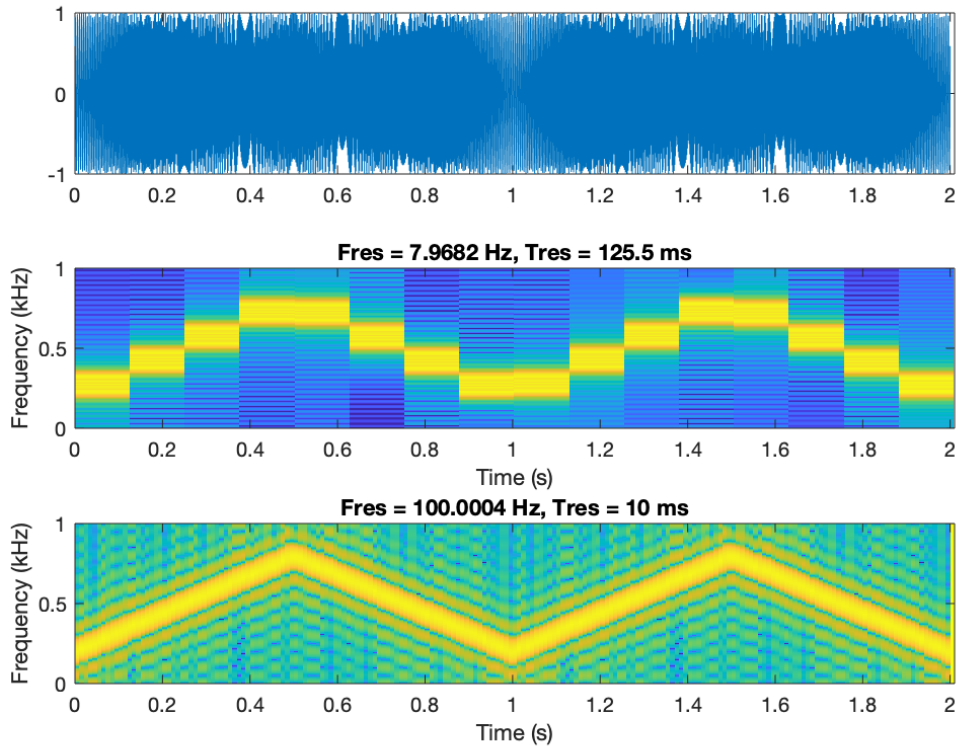
Şekil 2.17 : Bir ses sinyaline ait spektrogram örneği.



Şekil 2.18 : Bir ses sinyaline ait spektrogram örneği. (a) 2-boyutlu görünümü, (b) 3-boyutlu görünümü.

Grafik yakından incelenirse, düşük frekanslı bileşenlerin genlik değerlerinin, daha yüksek frekanslı bileşenlere göre daha yoğun olduğu gözlemlenecektir. Bu durum birçok çoklu ortam ses dosyasında bu şekildedir. Çünkü, perküsyon ve bas enstrümanları düşük frekans skalasında ses sinyali üretirler ve genellikle tüm ses verisi boyunca yer alırlar. Böylece, bu enstrümanlara ait düşük frekanslı sinyal bileşenleri, diğer enstrümanların düşük frekanslı bileşenleri ile üst üste gelerek bu duruma yol açarlar [75].

Yukarıda da belirtildiği üzere STFT'nin amacı, sinyali, hareket eden pencere tekniği yardımıyla, parçalar halinde üst üste işleme sokarak, bu parçaların DFT'ni elde etmektir. Bu sebeple STFT, kısa-terim işleme tekniği sınıfındadır. Burada, yukarıda bahsedildiği üzere, pencereleme işlemini gerçekleştiren hareketli pencere, verilen örnekleme frekansına bağlı olarak, frekans çözünürlüğünü belirler. Uzun seçilen pencerelerin frekans çözünürlükleri artarken, zaman çözünürlükleri azalacaktır. Tersi durumda ise, kısa tutulacak pencere boyutu, zaman çözünürlüğünü yükseltip daha detaylı sonuçlar verse de frekans çözünürlüğü bakımından zayıf sonuçlar doğuracaktır [57]. Bu durum Şekil 2.19 da görselleştirilmiştir.



Şekil 2.19 : Zaman ve frekans çözünürlüğü arasındaki ilişki [76].

Ses analizi uygulamalarında kısa-terim pencere uzunluğu genellikle 10–50 ms aralığında olması tercih edilir. MPEG7 ses standartı bu değerin 10 ms'nin katları olmasını önermektedir [77].

2.6 Ses Parmak İzi

Ses parmak izi yapısı akademi ve endüstri tarafından oldukça yoğun ilgi gören bir konudur. Bu alanda geliştirilen Shazam, SoundHound, Intonow, Viggie ve MusicID gibi uygulamalar bugün başarılı bir şekilde kullanılmakta olan uygulamalardır. Benzer şekilde akademik alanda ses parmak izi ile ilgili literatür incelendiğinde Google [4, 78], Philips [2, 79], Telefonica [5, 80], Microsoft [81] ve Gracenote [82] gibi şirketlerin çalışmaları öne çıkmaktadır [83].

Literatürde birçok çeşit müzik tanıma ve geri erişim (retrieval) metodu bulunmaktadır. Bu metodlardan en eskisi, dizi temelli erişim metodudur. Bu yöntemde, ses sinyalleri dizilere dönüştürülerek, bu diziler üzerinden işlemler gerçekleştirilmektedir [84].

Miller ve Baluja [2], [3] ve [8] 'de önerilen yöntemleri dikkate alarak, sesi parmak izlerine dönüştürmek için daha hızlı ve doğru çalışan bir yöntem önermişlerdir. Bu yöntem bir dalgacık tabanlı ses parmak izi algoritmasıdır. Önerilen sistemin, %2' ye kadar zaman ölçeği değişimleri ve %10'a kadar tempo değişimleri de dahil olmak üzere farklı bozulmaların varlığında dahi düzgün çalıştığı tespit edilmiştir [85].

Zhu, görüntüler için [7] nolu çalışmada önerilen Ölçekle Değişmeyen Özellik Dönüşümü'ne (Scale Invariant Feature Transform-SIFT) dayanan bir yerel ses özelliği çıkarma algoritmasını [9] nolu çalışmasında önermiştir. Ses özelliklerini çıkarmak için SIFT'yi ses sinyalinin logaritmik olarak ölçeklenmiş bir spektrogramına uygulamıştır. SIFT özellikleri çok yönlü yerel görüntü ölçeklendirmelerine karşı oldukça dayanıklıdır. Bununla birlikte, tempo değişimi spektrogramın zaman eksenini ölçeklendirir ancak frekans eksenini ölçeklendirmez, bu sebeple [6] 'da kullanılan yöntem tempo değişikliklerine karşı hassastır [85].

Serra ise, [10] nolu referansta ses sinyalinin zaman-kroma göstergesine dayanan global bir ses özelliği çıkarma algoritması önermiştir. Bu konuda kapsamlı deneyler yapmış ve bir ses sinyalinin zaman-kroma gösteriminin, ses saptama algoritmalarını tasarlamak için umut verici bir platform olduğunu göstermiştir [85].

Ses parmak izi, spektral tabanlı ses tanıma sistemlerinde oldukça sık kullanılan, ses verisine ait eşsiz bir kimlik oluşturan çok önemli bir tekniktir. Bu yöntem sayesinde ses verileri kompakt bir yapıda incelenebilir.

Ses parmak izi tasarlanması ve hesaplanması için oluşturulmuş çeşitli yöntemler mevcuttur. Bazı tür ses parmak izleri Mel-Frequency Cepstral Coefficients (MFCC) [88] yöntemi yardımıyla ses parmak izlerini oluştururlar. [89-94] nolu çalışmalar bu yöntemi baz alarak oluşturulmuşlardır.

Bir diğer tür ses parmak izi oluşturma yöntemi ise Bark ölçeği spektrogramları [2, 95] yardımıyla ses parmak izlerinin hesaplanmasıdır. [81] nolu çalışma bu yöntemi temel alarak ses parmak izlerini oluşturmuştur.

Bir başka tür ses parmak izi oluşturma yönteminde ise bir dizi düşük seviyeli tanımlayıcılardan faydalanılmıştır. [96] nolu çalışma bu yöntem ile gerçekleştirilmiştir.

[97] nolu çalışma çerçeve bazlı özellik vektörlerinden oluşan kısa zincirler halinde ses parmak izlerini oluşturur. Bu tarz yaklaşımlarda vektör kuantalama [97], eşikdeğer tekniği [95] veya geçici istatistikler [98] istenilen gürbüzlüğü elde etmek için kullanılmaktadır.

Diğer bir grup ses parmak izi yönteminde, spektral tepeler [3, 99] baz alınarak ses parmak izleri üretilmiştir. Bu alanda aralarında bu çalışmanın da temel aldığı Shazam algoritması yer almaktadır. [3, 5, 100-103] nolu çalışmalarda spektral tepeler yardımıyla, ses verilerinden ses parmak izleri oluşturulmuştur.

Bir diğer yöntemde [104], karakteristik dalgacık katsayıları gibi seyrek karakteristik noktalar temel alınmıştır.

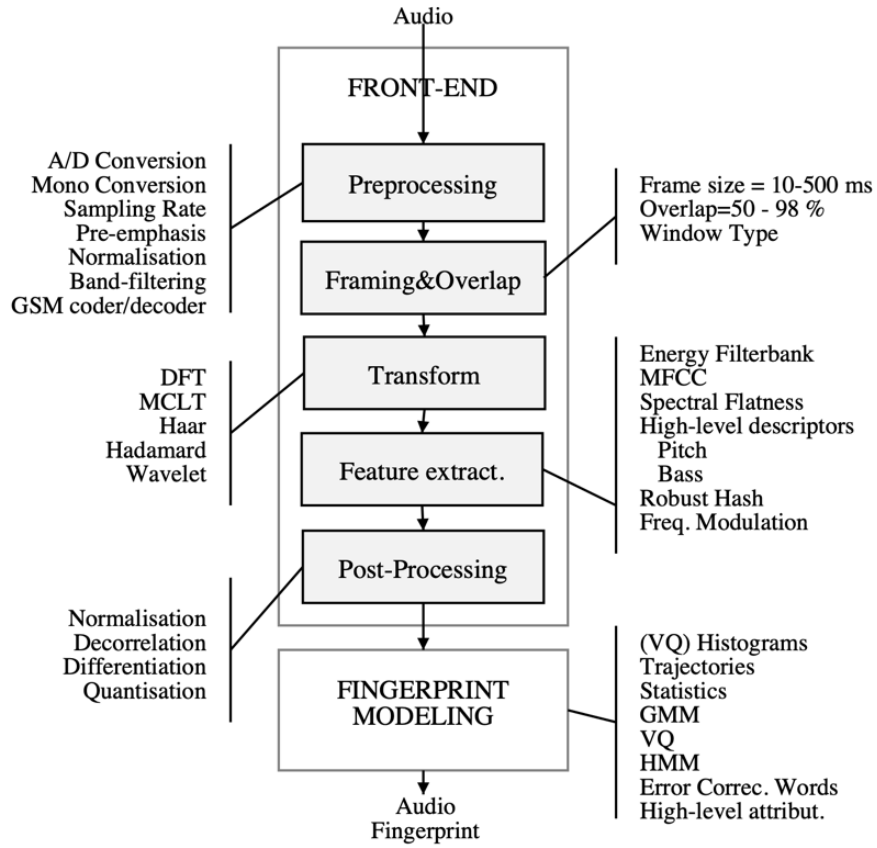
Ses parmak izi oluşturulması aşamasında en çok tercih edilen bir diğer yöntemde ise, Philips algortimasında kullandığı, spektruma ait altbant enerji değerleri baz alınarak ses parmak izlerinin üretilmesi tekniğidir. [2, 80, 105-111] nolu çalışmalar bu yöntemi temel alarak oluşturulmuştur.

Haitsma ve arkadaşları tarafından yapılan [2, 113] nolu çalışmalarda ses parmak izi şu şekilde tanımlanmıştır: Ses parmak izi, bir ses dosyasının kısa bir özetidir. Bu nedenle, bir ses parmak izi fonksiyonu F , büyük bir ses nesnesini, sınırlı boyuttaki bir parmak izine dönüştürür. Bu noktada kriptografideki hash fonksiyonlarıyla benzerlik kurulabilir. Kriptografik hash fonksiyonu büyük bir nesneyi, küçük bir hash değerine dönüştürür [86,

87, 112]. Bir kriptografik hash fonksiyonu iki büyük objenin eşit olup olmadığını anlamak için, sadece hash değerlerinin eşit olup olmadığına bakılarak bu durumun doğruluğuna karar verilmesine olanak sağlar. Hash değerlerinin matematiksel olarak eşit olması, çok küçük bir hata ihtimali ile, orijinal objelerin de eşit oldukları anlamına gelmektedir.

Kriptografik hash fonksiyonları ile bir nesnenin, verilen bir nesne kümesi içinde olup olmadığını anlamak kolaylaşır. Bütün veriyi saklamak ve karşılaştırmak yerine sadece hash değerlerinin saklanıp karşılaştırılmasıyla, o veri objesinin aranan kümede olup olmadığı anlaşılabilir. Bu açıklamaya göre, kriptografik hash fonksiyonlarının, ses parmak izi fonksiyonu olarak kullanılabileceği anlaşılabilir. Fakat bu tam olarak doğru değildir, çünkü parmak izi fonksiyonu, hash fonksiyonlarının sağladığı matematiksel eşitliği değil, algısal eşitliği sağlamalıdır. Örneğin bir ses dosyası farklı sıkıştırma teknikleriyle saklanabilir. Her bir sıkıştırma örneğine ait hash fonksiyonu yardımıyla üretilcek hash değeri, çok küçük bir olasılık dışında, farklı değerlerde olacaktır. Ama aynı şarkıyı farklı sıkıştırma teknikleriyle kaydedilmiş farklı dosyalardan çalıp dinlediğimizde bile aynı şarkı olduğunu anlarız. Bu yüzden parmak izi yöntemleri, bir ses verisini dinlediğimizde algıladığımız şekilde bir karşılaştırma yapabilmeyi sağlamalıdır.

Ses parmak izi oluşturmaya yönelik birçok yöntem olmasına rağmen, hepsinin ortak olarak paylaştığı bir şemayı Haitsma ve arkadaşları Şekil 2.20 de ki gibi bir diyagramla oluşturmuşlardır [1].



Şekil 2.20 : Ses parmak izi oluşturma şeması [1].

Bir başka kaynakta, ses parmak izi tekniği, ses verileri için eşsiz ve kompakt bir kimlik oluşturan ve bu kimlikleri birbirinden ayırt etmede faydalanılan bir yöntem olarak tanımlanmıştır. Bu yöntem ile ses dosyalarına ait içerikler herhangi bir meta veri veya filigran bilgisi olmadan denetlenebilir. İçerik tabanlı tanımlama teknolojileri (Content-Based Identification Technologies – CBID) olarak tanımlanan bu yöntem, eldeki ses verisine ait akustik parametreler ile işlemlerini gerçekleştirir. Bu yöntem, 90’lı yılların başında ilgi kazanan ve yapılan önemli çalışmalar ile, özellikle [114] nolu çalışma ile bir sıçrama yaşayan ve günümüzde hâlen kullanılmakta olan ses filigranlama (Audio Watermarking) yöntemi ile bazı önemli farklılıklara sahiptir. Ses filigranlama ses verilerine ait içeriği kontrol etmekte kullanılan bir yöntemdir. Ses filigranlamada, insan duyu sistemi dikkate alınarak, ses verisine, bu sesin algılanmasında herhangi bir bozulmaya yol açmayacak şekilde bir kod yerleştirilir. Daha sonra bu kodu tanımlayabilecek sistemler ile dosya takibi yapılmaya çalışılır. Bu yöntem aha çok yasal telif haklarının korunmasında kullanılan bir yöntemdir.

Ses parmak izi tekniğinde ise, mevcut ses verisi üzerinde herhangi bir değişiklik yapılmaz. Ses verisine ait temel yapıdan veriler türetilerek ses dosyalarının hızlı ve yüksek

doğruluk oranları ile tespit edilmesi amaçlanır. Bu yöntem ses filigranlamaya göre çok daha karmaşık işlemler gerektirir [116]. Ses filigranlama ile bir kaydın algısal olarak özdeş kopyaları arasında ayırım yapmak mümkünken, bu durum ses parmak izi tekniğinde gerçekleştirilemez.

Ses parmak izi, yukarıda bahsedildiği üzere çeşitli metodlar kullanılarak (MFCC, Bark ölçeği spektrogramı, spektral tepeler, vb.) oluşturulur. Ses parmak izi oluşturulduktan sonra, elde edilen bu veriler ile aynı yöntem ile oluşturulan veri tabanındaki veriler arasında yapılan karşılaştırma işlemlerine göre bir sonuç elde edilir. Bu işlem sırasında, sorgusu yapılan ses verisine ait ses parmak izlerinden oluşturulan veriler ile veri tabanında bulunan aynı yöntemler ile üretilmiş veriler karşılaştırılır. Bu noktada, karşılaştırmaya dahil edilen veri miktarı çok fazla olduğu için, hızlı ve efektif bir karşılaştırma algoritması oluşturmak oldukça önemlidir.

Haitsma ve arkadaşları [2] nolu çalışmada, bir ses parmak izi sisteminde, bir giriş sinyalini işleyerek bir grup karakteristik özellik oluşturmuş ve bu özellikleri parmak izi modelleme aşamasına aktarmışlardır. Oluşturulan bu karakteristik özellikler ile sistemin, kanal bozulmalarına ve gürültüye karşı dayanıklı olması amaçlanmıştır. Çalışmada bu kısım sistemin hazırlık aşaması olarak geçer ve beş aşamadan oluşur [1]:

- 1. Önileme:** Bu adımda, ses sinyali dijital hale getirilir ve kuantalanır. Ardından, eğer sinyal çok kanallı yapıdaysa, tek kanallı yapıya çevrilir. Son olarak, eğer örnekleme frekansı istenen değerde değilse, yeniden örnekleme ile istenen değerde olması sağlanır.
- 2. Çerçeveleme:** Çerçeveleme, ses sinyalinin bir pencere fonksiyonu (örn. Hanning) yardımıyla, eşit parçalar haline bölünmesi işlemidir. Bu işlem yapılırken, ses sinyalinin büyük bir kısmı baskılanır [117]. Çünkü pencere fonksiyonlarının sınır değerleri oldukça küçüktür. Bu durumda sinyalin enerjisi baskılanmış olur ve sinyale ait bilgiler gölgelenmiş olur. Bunun önüne geçmek için, çerçeveler üst üste bindirilerek işlemler gerçekleştirilir, yani “overlapping” tekniği ile azaltılmaya çalışılır.
- 3. Dönüşüm:** Bu adımda hedef, pencerelenmiş sinyalden, işlenecek kısımların açığa çıkarılması ve veri miktarının azaltılmasıdır. Bunun için, çoğunlukla FFT ile zaman domeninden, frekans domenine geçiş tercih edilir. Bunun yanında, daha başka birçok çeşit dönüşüm mekanizmaları da mevcuttur. DCT (Discrete Cosine Transform) [118], the Walsh-Hadamard dönüşümü [119], Modüle Edilmiş Karmaşık Dönüşüm (MCLT) [120], Tekil Değer Ayrışması [121] vb. bu dönüşümlere örnektir.

4. **Özellik çıkarma:** Dönüşümden sonra, zaman-frekans verisinden en son akustik özellikler çıkartılır. Buradaki amaç, boyutluluğu azaltmak ve bozulmalara karşı güçlü bir yapı açığa çıkartmaktır. Bu konuda birçok teknik önerilmiştir. Bunlardan bazıları; MFCC (Mel-Frequency Cepstrum Coefficients) [88], Spectral Flatness Measure [96], band representative vectors [122] vb. dir.
5. **Rötuş:** Ses sinyalinin geçici değişimlerini yakalamak için bazen zamanın yüksek dereceli türevleri gerekmektedir. Örneğin, [88] nolu çalışmada, dördüncü adımda açığa çıkarılan MFCC özelliklerinin yanısıra, oluşturulan son özellik vektörü, özelliğe ait türevleri ve ivmelenmeleri, enerjinin türevleri ve ivmelenmeleri kadar içermektedir. Özelliklerin türevi gürültüyü artıracak olsa da ortaya çıkan bozulmalar, doğrusal-zamanla değişmeyen bir filtre kullanılarak azaltılabilir [123].

Cano ve arkadaşları ses parmak izi tekniği ile ilgi yapmış oldukları [116] nolu kapsamlı çalışmada, bir ses parmak izi tekniğinin sahip olması gereken özellikleri ve bu özelliklerin önemini belirtmişlerdir. Bu çalışmalarında, farklı ses parmak izi teknolojilerini değerlendirmek ve karşılaştırmak için, [124] nolu çalışmada ortaya konulan, IFPI (Uluslararası Fonografik Endüstrisi Federasyonu) ve RIAA (Amerika Kayıt Endüstrisi Birliği) tarafından oluşturulmuş, tanımlama sistemlerinin değerlendirilmesine yönelik çalışmalardan faydalanmışlardır. Bu çalışmaya göre, bu tür sistemler en başta hesaplama açısından verimli ve sağlam olmalıdır. Devamında beklenen nitelikler şu şekilde aktarılmıştır:

- **Doğruluk:** Doğru, eksik ve hatalı (yani yanlış pozitif) sayıları ifade eder.
- **Güvenilirlik:** Telif haklarını koruma kapsamında, sorgusu yapılacak ses dosyasının bu kontrolleri yapan şirketlerin kontrol listesinde mevcut olup olmaması, yasal işlemler açısından oldukça büyük bir önem arz etmektedir. Bu sebeple oluşturulan ses parmak izi temelli yöntemin güvenilir olması istenir. Bunun için de sistemin yanlış pozitif sonuçlar vermemesi gerekir. Bu konuda [125] nolu çalışmada yazarlar bu durum üzerine çalışmışlardır.
- **Sağlamlık:** Sıkıştırma seviyesi, arka plan gürültüsü, D/A-A/D (Dijital/Analog-Analog/Dijital) dönüşüm işlemleri, ses kodlayıcılar veya iletim kanalında meydana gelebilecek her türlü olumsuz duruma karşılık doğru tespiti yapabilme yeteneğidir.
- **Güvenlik:** Çözüm sisteminin her türlü yanıltıcı hamlelere karşı dayanıklılığını temsil eder. Bu yanıltıcı hamleler, parmak izi tanımlama algoritmalarını adlatmaya yönelik hazırlanmışlardır.

- **Çok Yönlülük:** Ses formatına bakılmaksızın sesi tanımlama yeteneğidir. Aynı veri tabanını farklı uygulamalar için kullanabilmeyi de ifade eder.
- **Ölçeklenebilirlik:** Çok büyük boyuttaki veri tabanları veya çok sayıda eşzamanlı tanımlama durumlarında gösterilen performansı ifade eder. Bu durum, sistemin doğruluğunu ve karmaşıklığını etkiler.
- **Karmaşıklık:** Parmak izi çıkarmanın hesaplama maliyetlerini, parmak izinin boyutunu, aramanın karmaşıklığını, parmak izi karşılaştırmasının karmaşıklığını, veri tabanına yeni öğeler eklemenin maliyetini, vb. durumları ifade eder.
- **Kırılganlık:** İçerik bütünlüğü doğrulama sistemleri (CBID) gibi bazı uygulamalar, içerikteki değişikliklerin tespit edilmesini gerektirebilir. Bu, sağlamlık gerekliliğine aykırıdır, çünkü parmak izinin içeriği koruyan dönüşümlere karşı sağlam olması gerekir, ancak diğer bozulmalara karşı sağlam olması gerekmez.

Bu durumların yanı sıra, bir ses parmak izi şu önemli özellikleri de barındırmalıdır [116]:

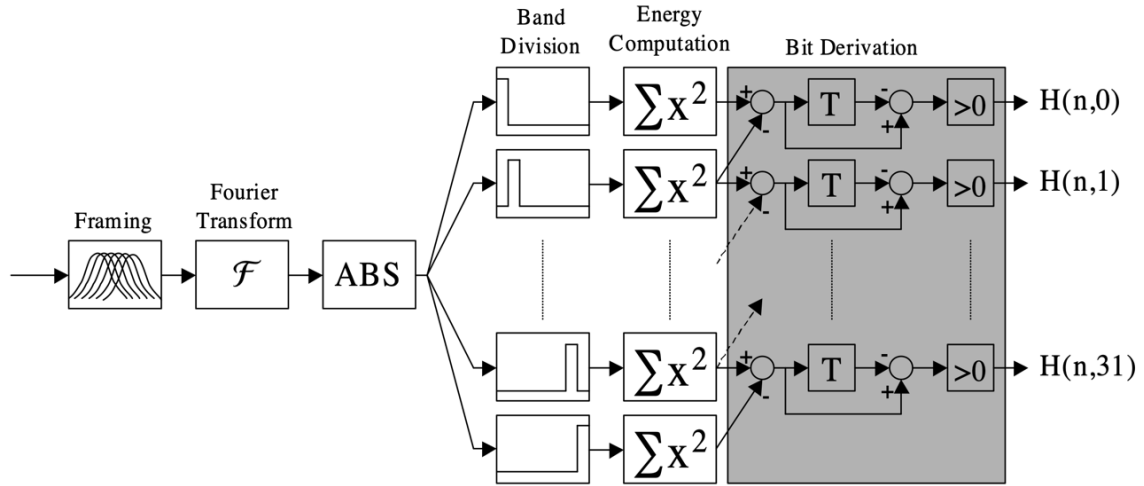
- Kaydın algısal özeti iyi belirlenmelidir. Parmak izi, akustik bakımdan ilgili maksimum bilgiyi tutmalıdır. Bu özet yapıdaki bilgi, çok sayıda parmak izi üzerinde ayırım yapılmasına izin vermelidir. Bu, karmaşıklık ve sağlamlık gibi diğer gereksinimlerle çelişiyor olabilir.
- Bozulmalara karşı değişmez olmalıdır. Bu durum, sağlamlık gerekliliğinden kaynaklanmaktadır. Bununla birlikte, içerik bütünlüğü uygulamaları, kasıtlı manipülasyonları tespit etmek ve içeriği koruyan bozulmalar için bu kısıtlamayı gevşetir.
- Kompakt yapıda olmalıdır. Küçük boyutlu bir temsil, karmaşıklık açısından ilginçtir, çünkü çok sayıda (belki milyonlarca) parmak izinin depolanması ve karşılaştırılması gerekir. Bununla birlikte, aşırı derecede kısa bir sunum, kayıtlar arasında ayırım yapmak için yeterli olmayabilir, bu da doğruluğu, güvenilirliği ve sağlamlığı etkiler.
- Kolayca hesaplanabilir yapıda olmalıdır. Karmaşıklık nedenlerinden dolayı, parmak izinin çıkarılması aşırı derecede zaman alıcı olmamalıdır.

Ses parmak izi tekniği temelde dört kullanım alanı ile karşımıza çıkmaktadır [116]. Bunlar; ses tanımlama uygulamaları, bütünlük doğrulama uygulamaları, filigranlama işleminde yardımcı uygulama olarak kullanımı ve içerik tabanlı ses tanımla ve işleme uygulamalarıdır. Bu çalışmada bir ses tanımlama uygulaması için kullanılmıştır.

Daha önce ses parmak izi tekniğine ait önemli iki uygulamanın Philips ve Shazam teknikleri olduğu belirtilmişti. Bu tekniklere ait temel bilgiler şu şekildedir:

2.6.1 Philips tekniği

Şekil 2.2 Philips Research firmasının yayınlamış olduğu, kendi adıyla anılan Philips ses parmak izi algoritmasına ait sistemin blok diyagramını göstermektedir [79].



Şekil 2.21 : Philips ses parmak izi sistemine ait blok diyagram.

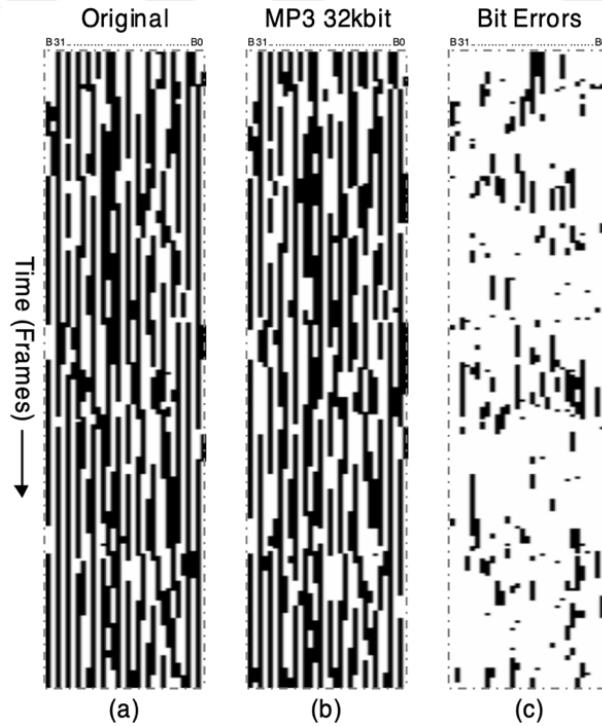
Diğer ses parmak izi oluşturma algoritmaları gibi, bu algoritma da iki ana kısımdan oluşmaktadır. İlk kısım ses parmak izini meydana getiren kısımdır. İkinci kısım ise çıkarılan bu ses parmak izlerini, ses parmak izi veri tabanında sorgulanmasını sağlayan kısımdır.

Philips ses parmak izi algoritmasında, ses verileri 11.6 ms lik uzunluklara karşılık 32 bitlik ses alt-parmak izleri oluşturulacak şeklinde işlenir. Bunun için ses verisi 0.37 ms lik uzunluklar halinde Hanning penceresi yardımıyla pencerelenerek işleme dahil edilir. Hassas bir inceleme işlemi için bu çerçevelere 31/32 oranında üst üste binme oranı uygulanır. Bu durum oluşan parmak izlerinin uzun ve yavaş değişir formda olmasını sağlayarak başarı oranını artırır. Bu üst üste binme oranı ile ses verisine ait her 11.6 ms için bir kodlama oluşturulmuş olunur. Uygulanan bu çerçeveleme işleminin ardından ses verisinin frekans spektrumu hesaplanır. Bunun için Fourier dönüşümden yararlanılır. Her bir çerçeve için 32 bitlik ses parmak izi, tahmini spektral güç yoğunluğunun 33 adet üst üste binmeyen frekans bandından türetilir. Bu frekans bandı 300 Hz ile 2000 Hz arasındaki değerlerin logaritmik olarak hizalandırılmasıyla oluşturulur. Bu durum insan duyu sistemi ile paralellik göstermektedir. İnsan duyu sistemi de bu tarz bir düzende çalışmaktadır. Bu frekans bantları

belirlendikten sonra, bu bantlar arasında kalan frekans bant enerjileri farkı hesaplanır. Hesaplanan bu değerler hash değerlerinin hesaplanmasında kullanılır. 33 farklı banttan elde edilen 32 adet enerji fark değeri, 32 bitlik bir parmak izini meydana getirir. Bant enerjileri yardımıyla hesaplanan bu değerlere ait eşitlik denklem 2.10 da verilmiştir: [79].

$$F(n,m) = \begin{cases} 1 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) > 0 \\ 0 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) \leq 0 \end{cases} \quad (2.10)$$

Burada $F(n, m)$, n . çerçevenin m . enerji bandını, $E(n, m)$ ise bu n . çerçeveye ait hesaplanan hash değerinin m . bitini temsil etmektedir. Bu yöntemle oluşturulmuş hash değerlerine ait yapı Şekil 2.22 de verilmiştir. Burada yaklaşık 3 sn lik bir ses verisinden, yukarıdaki eşitlik baz alınarak 256 alt çerçeve halinde hash değerleri, beyaz pikseller 1'i, siyah pikseller ise 0'ı temsil edecek şekilde oluşturulmuştur. Şekil 2.22 (a) ve Şekil 2.22 (b) bir ses verisinin sırasıyla orijinal ve 32 Kbps örnekleme hızında MP3 (MPEG-1 Audio Layer III) formatında sıkıştırılmasıyla oluşturulmuş formatlarından üretilmiştir. Aradaki hata oranı sıkıştırma işleminden kaynaklı ses verisinde meydana gelen veri değişikliklerinden kaynaklıdır [79].



Şekil 2.22 : Philips ses parmak izi algoritması ile oluşturulmuş (a) Orijinal müziğe ait oluşturulmuş hash değerleri, (b) sıkıştırılmış ses dosyasına ait üretilmiş hash değerleri, (c) a ve b şekilleri arasındaki bit hataları (Siyah bölgeler) (BER=0.115-Bit Error Rate) [79].

Bu sistem ile bir ses verisinin 3 saniyelik bir parçası ile 256 adet alt parmak izi 8192 bitten oluşturularak oldukça büyük veritabanlarından başarılı bir şekilde ses verisinin tespiti yapılabilmektedir.

2.6.2 Shazam tekniği

Shazam, temelleri 1999 yılında Birleşik Devletler Berkeley’de bulunan Kaliforniya Üniversitesi’nde atılan, günümüzün en popüler müzik arama uygulamalarından birisidir. Shazam firmasının 2019 yılında paylaşmış olduğu bilgilere göre, aylık kullanıcı sayısı yaklaşık 150 milyon kişi iken, günlük yapılan arama sayısı yaklaşık 20 milyon adettir [126].

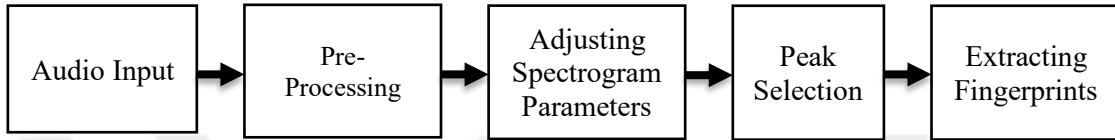
Shazam algoritmasının ardındaki en özgün yapı, ses sinyalinin elde edilen spektrogramın basit koordinatlar şeklinde sadeleştirilmesidir. Bu işlem neticesinde elde edilen yapıya parmak izi adı verilir. Spektrogram yapısı üzerinden işlemlerin yürütülmesi en önemli nedeni, spektrogramların sunmuş oldukları gürbüz yapı ve bünyesindeki bileşenler üzerinde oldukça kolay bir şekilde işlemlerin yürütülebilmesinden dolayıdır. Bu avantaj, özellikle parmak izi elde edilirken birçok kolaylık sağlar. Spektrogram içerisindeki tepe noktalar üzerinden işlem yapabilme imkanıyla birlikte, ses sinyali işleme alanındaki en büyük handikaplardan birisi olan gürültü etkenine karşı büyük bir avantaj sağlanmıştır. Çünkü, bu noktalar gürültüden en az etkilenecek sinyal hakkındaki doğru bilgileri saklarlar [127]. Bu noktalar, spektrogram üzerindeki noktalar kümesinden, birçok farklı yöntemle açığa çıkartılabilirler. Çalışmanın temel aldığı bu teknik, metodoloji bölümünde tüm detaylarıyla paylaşılmıştır.

3. METODOLOJİ

3.1 Ses Parmak İzlerinin Üretilmesi

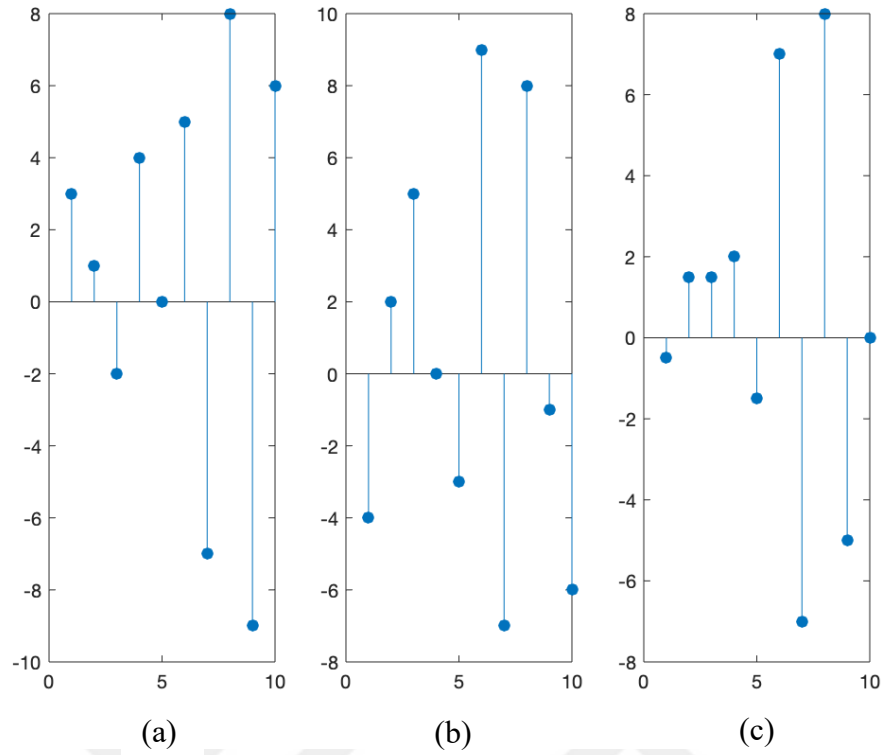
Shazam algoritması baz alınarak gerçekleştirilen bu çalışmada aşağıda sıralanan yöntemler ile çoklu ortam ses dosyalarının tespitine yönelik bir algoritma oluşturulmuştur.

Çalışmada, ses parmak izlerinin üretilmesi için izlenen yöneme ait akış diyagramı Şekil 3.1 de gösterilmiştir.



Şekil 3.1 : Ses parmak izi üretimine ait çalışmada izlenen yöntemin blok diyagramı.

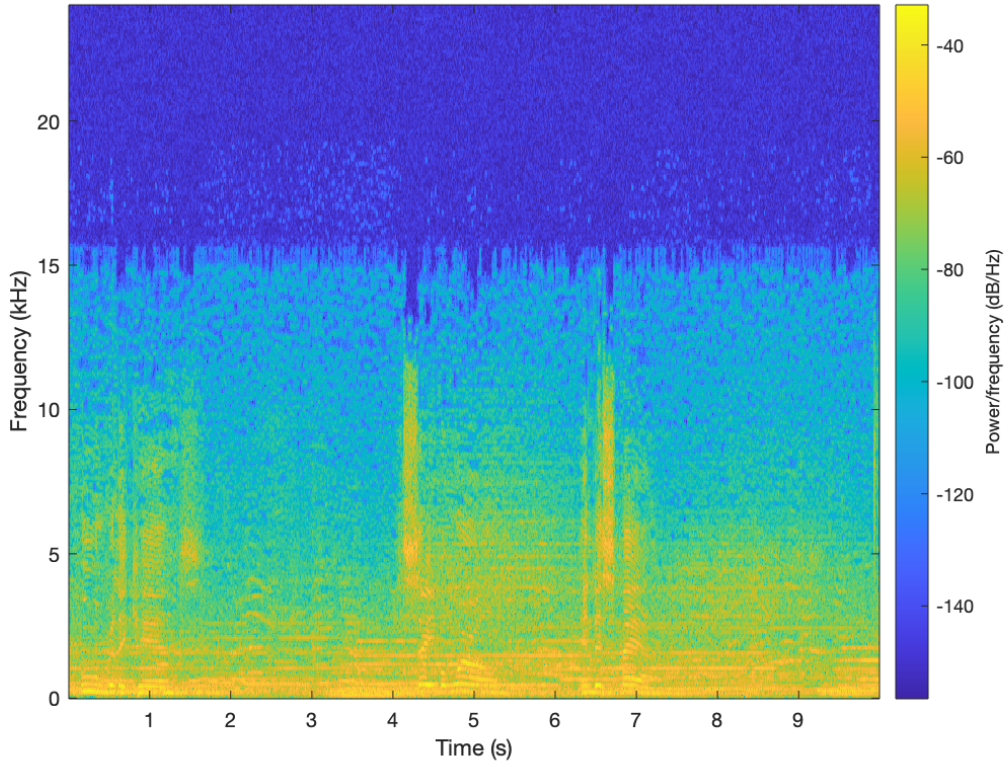
Öncelikle üzerinde işlem yapılacak çoklu ortam ses dosyasının kaç kanallı olduğu kontrol edilir. Bu noktada, tasarlanan algoritma için, işlenecek verinin tek kanallı (mono) yapıda olması gerekmektedir. Giriş ses sinyali verisi bu amaç doğrultusunda kontrol edilir, yapısı stereo formda, yani iki kanallı yapıda ise tek kanallı forma dönüştürülür. Stereo formdan mono forma dönüştürme işlemi için uygulanabilecek çeşitli yöntemler mevcuttur. Bu çalışmada tercih edilen yöntem, iki kanalın ortalama değerinin hesaplanması olarak tercih edilmiştir. Bu teknik ile algoritmanın iş yükünün mümkün olduğu kadar azaltılması amaçlanmıştır. Çünkü bu tarz uygulamalarda işlem sürelerinin minimum düzeyde olması çok büyük bir önem arz etmektedir. Şekil 3.2 de iki kanallı bir çoklu ortam ses dosyası parçasına ait sağ kanal (a) ile sol kanalın (b) aritmetik ortalaması alınarak elde edilen tek kanallı yeni ses verisi (c) basit bir örnek ile gösterilmiştir.



Şekil 3.2 : İki kanallı ses verisinden tek kanallı ses verisinin elde edilmesi. Yatay eksen zamanı, düşey eksen ise genliği ifade etmektedir.

Bir sonraki adımda, eldeki tek kanallı çoklu ortam ses verisinin örnekleme hızının önceden belirlenen değerde olması sağlanır. Çoklu ortam ses dosyaları frekans domeninde incelendiğinde, bu alandaki verilerin 0-4 kHz frekans bant aralığındaki genlik değerlerinin, bu bant aralığı dışındaki genlik değerlerine göre daha yüksek olduğu görülür.

Şekil 3.3, 10 saniye uzunluğunda bir çoklu ortam ses dosyasına ait spektrogramı göstermektedir. Yukarıda bahsedilen durum, bu görsel üzerinde kolaylıkla gözlemlenebilmektedir. Sarı renk tonunun en yoğun bulunduğu bölgeler bu frekans değerleri arasında kalmaktadır.



Şekil 3.3 : Veri tabanından rastgele seçilmiş bir ses dosyasının 10 saniyelik bir bölümüne ait spektrogram görseli.

Bu durumun temel sebebi daha öncede bahsedildiği üzere, bu dosyalar oluşturulurken dikkate alınan ses kaynaklarının, ağırlıklı olarak nispeten daha düşük frekans bant aralığında ses sinyalleri üretmesinden dolayıdır [75]. Bu noktada 8 kHz lik yeni örnekleme frekansı ile ses dosyasının yeniden örneklenmesi tercih edilmiştir. Bu işlemle birlikte 0-4 kHz bant aralığı dışındaki veriler filtrelenmiş olur. Bu işlem alt örnekleme olarak adlandırılır. Örneğin 44.1 kHz örnekleme oranına sahip bir sinyal için 11.025 kHz örnekleme oranında olmasını sağlamak için 44.1 kHz’lik sinyalin bileşenleri sırası ile dörderli guruplar halinde alınır. Devamında bu örneklerin genliklerinin aritmetik ortalaması alınır ve elde edilen bu değer 11.025 kHz’lik yeni sinyalin bir bileşeni olarak atanır. Bu durum Şekil 3.4 te gösterilmiştir.



Şekil 3.4 : Alt örnekleme işlemi.

Önceki örnekleme değerindeki ses verisinden alınan dört bileşenin genlik değerlerinin (mavi çizgiler) aritmetik ortalaması ile yeni örnekleme değerine sahip ses verisinin bir bileşeni (kırmızı çizgi) elde edilir. Bu çalışmada alt örnekleme işlemi için Matlab'a ait `resample()` fonksiyonu kullanılmıştır.

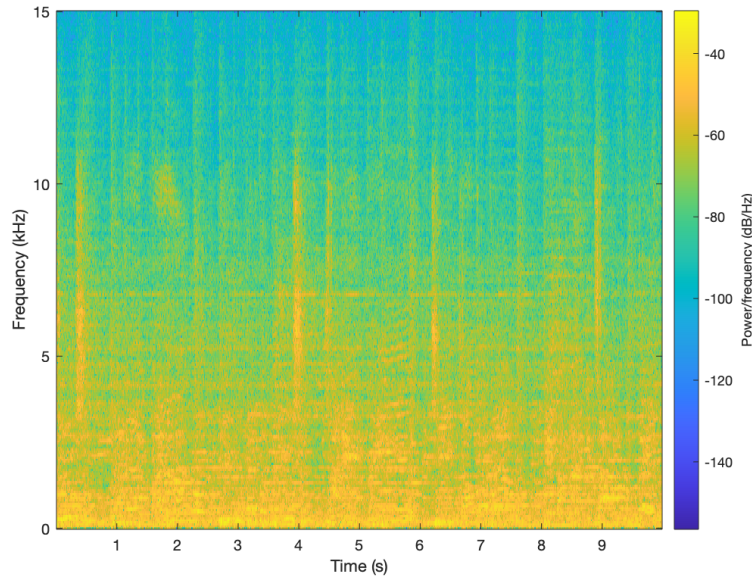
Uygulanan bu işlemlerden sonra, Fourier dönüşümü kullanarak zaman-frekans domenindeki bilgilerin elde edilip, spektrogramın oluşturulmasına sıra gelir. Ses sinyali işleme alanında, sinyallerin zaman-frekans domenindeki özelliklerinin gözlemlenmesinde, sağlamış olduğu birçok avantaj sebebiyle spektrogram yapısının sıklıkla kullanılan oldukça güçlü bir analiz aracı olduğu bir önceki bölümde sebepleriyle birlikte vurgulanmıştır. İşlenecek ses sinyali ham bir şekilde, doğrudan ele alınıp değerlendirildiği takdirde, bu sinyale ait genlik ve zaman değerleri üzerinden işlemler sürdürülecektir. Ancak bu metod sinyal hakkında çok az bilgi sunacaktır. Benzer şekilde, ses verisine Fourier dönüşümü uygulanırsa, bu defa sinyale ait frekans spektrumu ve bu spektruma ait genlik değerleri elde edilecek ancak zaman domenindeki değişimi gözlemlenemeyecektir. Bu durumun giderilmesi, sinyaldeki anlık değişimlerin önemli olduğu çalışmalarda olmazsa olmazdır. Bu noktada sinyale ait hesaplanacak spektrogram, durağan olmayan konuşma [128], biyomedikal [129], müzik [130], sismik sinyaller [131] gibi sinyallerin, aynı anda hem zaman hemde frekans domeninde incelenmesine olanak sağlayacaktır.

Bir sinyale ait frekans bilgisi Fourier dönüşüm yardımıyla elde edilebilir [132]. Ancak bu işlemle birlikte sinyale ait zaman bilgisi ortadan kalkmaktadır [133, 134]. Bu durum ilk olarak [135] nolu çalışmada vurgulanmış ve bu durumu çözmek için STFT (Short Time Fourier Transform – Kısa Süreli Fourier Dönüşümü) dönüşüm tekniği, bilinen diğer adıyla Gabor dönüşümü geliştirilmiştir. Bu dönüşüme ait eşitlik Denklem 3.1 de verilmiştir [56].

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x[n] \cdot w[n - mR] e^{-j\omega n} \quad (3.1)$$

Burada $w(n)$ pencere fonksiyonudur ve bu fonksiyon, dönüşüme zaman değişkenini dahil eden parametredir. Pencere fonksiyonları önceki bölümde bahsedildiği üzere, seçilen aralık dışında sıfır değerine sahip ağırlık fonksiyonlarıydı. Bu fonksiyonlar, üzerinde işlem yapılacak sinyalin parçalar halinde işleme alınmasına olanak sağlarlar. STFT uygulanırken bu fonksiyon yardımıyla seçilen sinyal parçasına Fourier dönüşümü uygulanır ve işlem sonunda pencere fonksiyonu merkez noktasından itibaren belirlenen ölçüde (R) kaydırılır. Sinyalin tamamına bu işlem uygulanana kadar bu işlemler sürdürülür. İşlem sonunda elde edilen matrise ait değerlerin genliklerinin kareleri spektrogramı oluşturur [136].

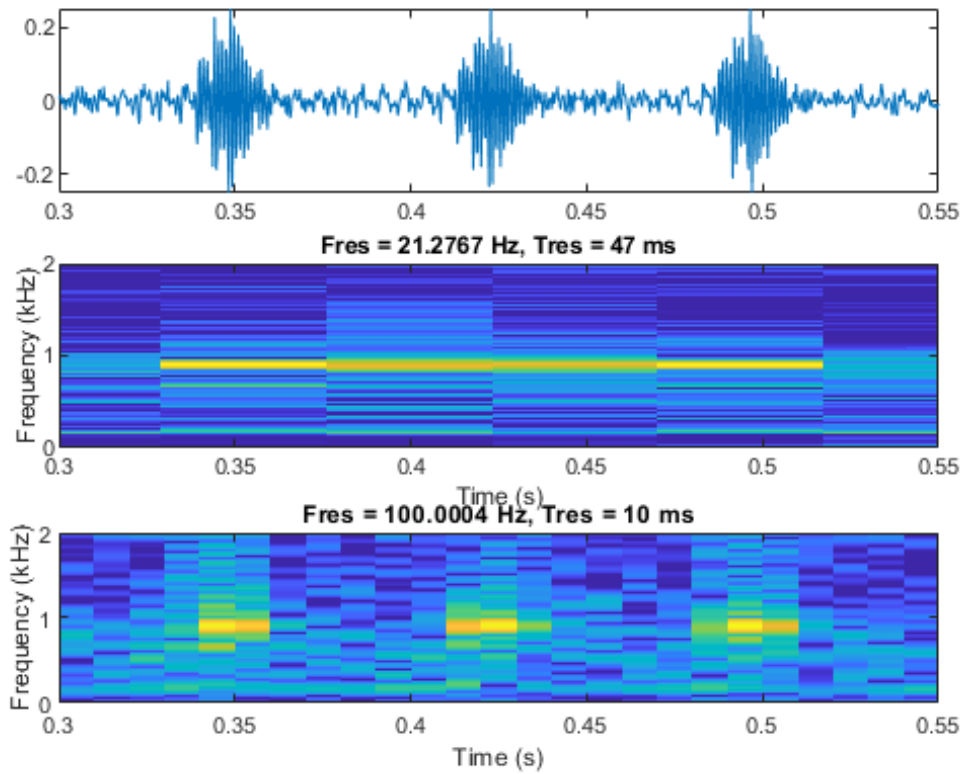
$$\text{spectrogram}(t, \omega) = |\text{STFT}(t, \omega)|^2 \quad (3.1)$$



Şekil 3.5 : Bir ses sinyaline ait spektrogram örneği.

Bir önceki bölümde belirtildiği üzere, STFT'nin en büyük dezavantajı uygun zaman-frekans çözünürlüğünün belirlenmesi problemidir. Bunun için spektrogram parametreleri dikkatli seçilmelidir. Pencere fonksiyonunun uzun seçilmesi frekans çözünürlüğü noktasında iyi sonuçlar verse de zaman çözünürlüğü konusunda belirsizlikler yaratacaktır. Benzer şekilde, kısa seçilen pencere uzunluğu zaman çözünürlüğü noktasında olumlu sonuçlar verir, fakat frekans çözünürlüğü açısından belirsizlikler yaratmaktadır [137]. Bu durum literatürde “belirsizlik ilkesi” olarak tanımlanmaktadır [138]. Çalışmanın uygulama adımlarında bu kavram ile ilgili testlere yer verilecektir.

Zaman ve frekans çözünürlüğü kavramları Şekil 3.6 da görselleştirilmiştir.



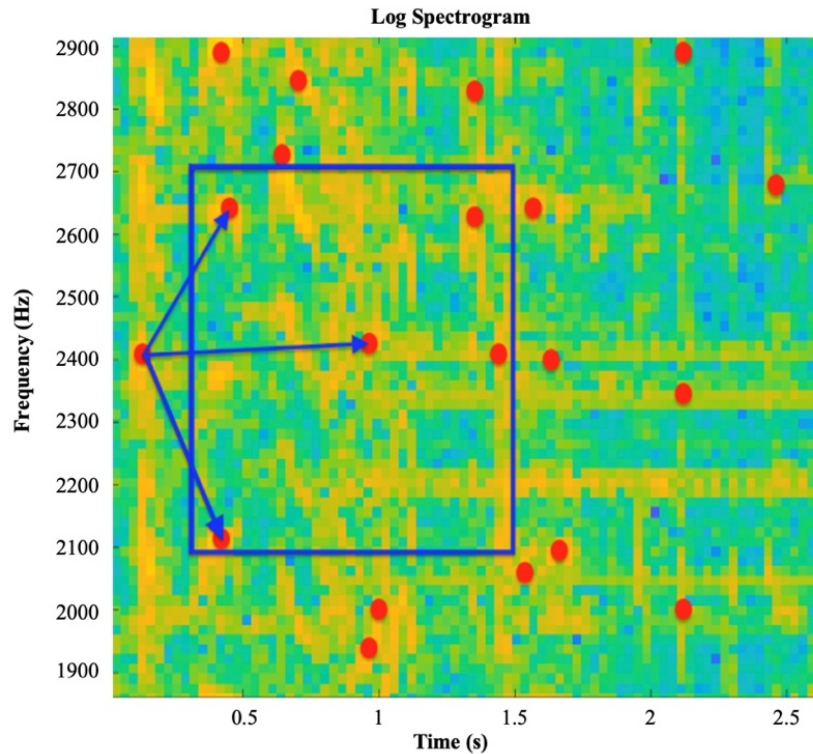
Şekil 3.6 : Zaman-frekans çözünürlüğünün spektrogram üzerindeki etkisi [139].

Konuşma algılama alanında 20-40 milisaniyelik pencere uzunluğu, bu tür ses sinyallerinin analizinde en iyi zaman-frekans çözünürlüğünü sunarken [140], müzik dosyaları için ise 64 milisaniyelik pencere uzunluğu iyi bir zaman-frekans çözünürlüğü sunmaktadır [130].

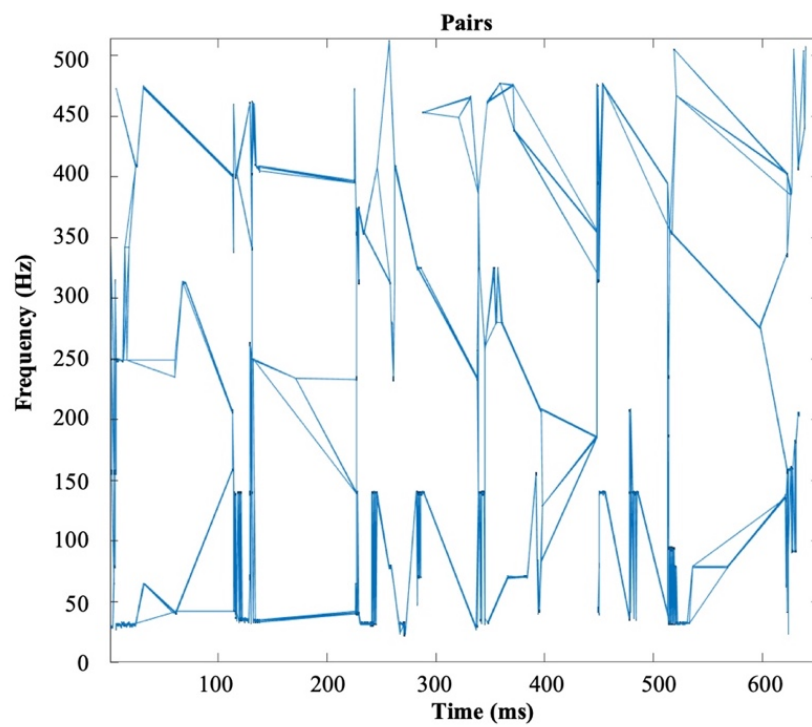
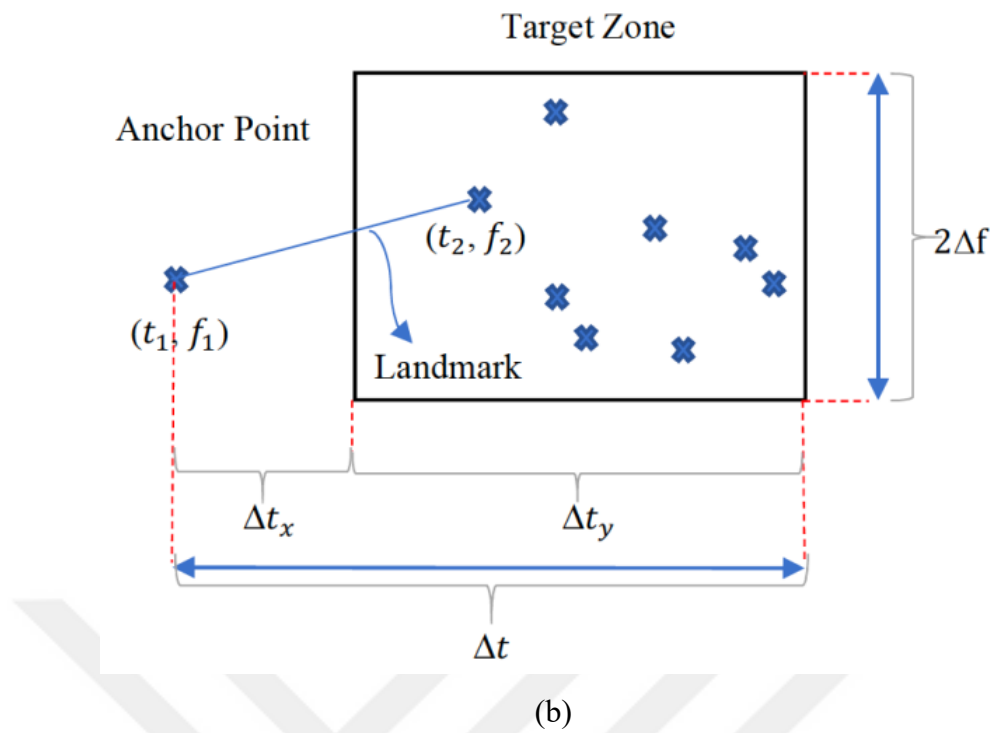
Spektrogram elde edildikten sonraki adımda, Wang'ın [3] nolu referansta önerdiği üzere, oluşturulan spektrogram üzerinde yüksek enerji yoğunluğuna sahip belirli noktalar

dikkate alınarak, zaman frekans ekseninde sadeleştirilmiş yeni bir yüksek enerjili noktalar haritası (constellation map) oluşturulmasındadır. Bu yeni harita oluşturulurken genlik verisi elimine edilir. Oluşturulan yeni harita veri tabanındaki kayıtlı haritalarla karşılaştırıldığında, yeni oluşturulan haritanın en yüksek oranda eşleştiği haritaya göre çoklu ortam ses dosyasının tespiti yapılabilir. Ancak oluşturulan yeni haritadaki yüksek enerji noktaları, bu tarz bir eşleştirme yöntemi için yüksek miktarda işlem yapılmasını gerektirdiğinden, bu eşleştirme uzun süre alabilir ve bu durum süreci yavaşlatır. Bu problemin üstesinden gelebilmek ve süreci hızlandırmak için Wang, kombinasyonel bölüntüleme (Hashing) yöntemini bu sürece dahil ederek çok büyük bir zaman ve işlem tasarrufu sağlamıştır.

Bu yöntemde her bir yüksek enerji noktası, ayrı birer temel nokta (anchor point) olarak ele alınır ve her bir nokta için belirli sayıda hedef nokta içeren bir hedef bölge hesaplanarak her bir temel nokta, sırasıyla hedef bölge içerisindeki noktalarla belirlenen sayılarda eşleştirilir. Bu eşleştirme bilgisi eşleştirilen temel nokta ve hedef noktasının ayrı ayrı zaman farkı ve frekans bilgilerini içerir ve bu bilgilerden hash değerleri türetilir. Her bir temel noktaya ait hash değeri, giriş dosyasının başlangıcından belirlenen temel noktaya kadar olan zaman bilgisini bir sayıyla gösterir. Her bir hedef bölge içerisindeki hedef nokta sayısı Wang'ın paylaştığı çalışmada belirtilmemiştir. Bu çalışmada bu değer 3 olarak belirlenmiş ve uygulamalar bu değere göre gerçekleştirmiştir.



(a)



hs	ts	SongID	hc	tc
648348	1	1	267082	1
648407	1	1	725896	1
274200	2	1	109211	3
536369	2	1	174666	3
268337	6	1	105290	4
333896	6	1	564104	4
.	.	.	.	.
.	.	.	.	.
≈	≈	≈	≈	≈
.	.	.	.	.
.	.	.	.	.
454823	2162	18	1376209	154
520371	2162	18	1376244	154
1819	2163	18	271189	156
263970	2163	18	664352	156
.	.	.	.	.
.	.	.	.	.
≈	≈	≈	≈	≈
.	.	.	.	.
.	.	.	.	.

(d)

Şekil 3.7 : (a) Hedef bölge şeması. (b)-(c) Seçkin noktaların eşleştirilmesi (Landmarks) (d) Üretilen ses parmak izi verilerinin kaydedildiği tablo yapısı. Soldaki tablo ses parmak izi veri tabanına, sağdaki ise sorgusu yapılan ses verisine ait oluşturulan ses parmak izi verilerinin kaydedildiği tablodur. Sorgulama işlemleri bu iki tablo arasında gerçekleştirilir.

Şekil 3.7, özetle bir spektrogramdan üretilcek parmak izi verilerinin hangi sırada oluşturulacağını göstermektedir. Şekil 3.7 (a) da spektrogram üzerinde seçilmiş bir hedef bölge (mavi çerçeve) ve bu bölge içerisinde kalan seçkin noktalar (kırmızı noktalar) gösterilmektedir. Şekildende görüleceği üzere her bir seçkin nokta, kendisinden sonra gelen üç farklı seçkin nokta ile ilişkilendirilmiştir. Şekil 3.7 (b) de ise bu durum bir grafik halinde verilmiştir. Bu şekilde iki seçkin nokta arasında oluşturulacak her bir ilişkilendirme “landmark” olarak adlandırılmıştır [141]. Şekil 3.7 (c) de ise landmark bağlantılarına ait oluşturulmuş bir örnek görsel verilmiştir. Şekil 3.7 (d) ise bu bağlantılardan türetilen, parmak izi verilerinin kayıt edildiği veri tablolarını göstermektedir. Şekilde solda bulunan üç sütunlu ilk tablo ses parmak izlerinin önceden hesaplanıp kaydedildiği veri tabanına ait tablodur. Burada hs, Song ID isimli çoklu ortam ses dosyasına ait, dosyanın ts anında üretilmiş hash değeridir. Sağda bulunan veri tablosu ise sorgusu yapılan çoklu ortam ses dosyasına aittir. Bu tablo soldaki tablodan farklı olarak Song ID sütunu içermez. Her iki tablodaki hash verileri birebir aynı metod ile üretilir. Yine burada hc değeri bu dosyaya ait tc anında üretilen bir hash değerini temsil eder. Sorgulama işlemleri bu iki tablo arasında oluşturulacak

karşılaştırma algoritmasına göre gerçekleştirilecektir. Benzerlik kriterini sağlayan veri tabanı dosyası bu işlemlerden sonra sorgusu yapılan dosya olarak belirlenecektir.

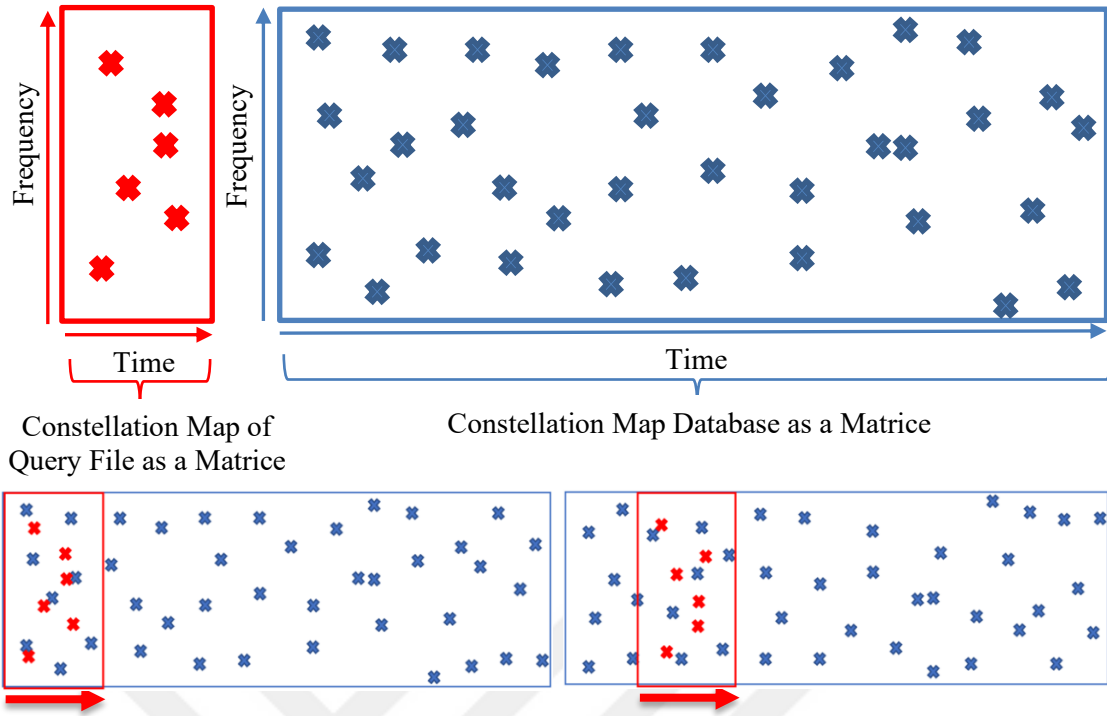
Hashing işlemi, M uzunluğundaki ikili değerlerin verilerini N uzunluğunda bir ikili değere dönüştüren bir işlemdir. Hashing işleminin bilinen en popüler ve yaygın kullanım örneği, bir ağ üzerinden dosya paylaşımı esnasındaki kullanım amacıdır. Dosya paylaşımı sırasında paylaşılan dosyadan bir hash değeri hesaplanır ve ardından gönderim işlemi gerçekleştirilir. İletim sırasında herhangi bir sorun yaşanmadığı takdirde, alıcı tarafında aynı metod ile üretilen hash değeri bu değer ile uyuşacaktır. Bu noktada ortaya çıkabilecek bir sorun, farklı dosyaların aynı hash değerini oluşturabilmesidir. Literatürde bu duruma "çarpmışma" (collision) denir. İstenmeyen bu durumun önüne geçmek için yeni teknikler geliştirilmektedir [79].

Bu çalışmada üretilen hash değerleri için çiftler halinde spektral tepeler kullanılmıştır. Her iki tepeden frekans ve zaman değerleri, [141] nolu referansta olduğu gibi bir hash fonksiyonu yardımıyla hash değerlerine dönüştürülür. Bu fonksiyon Denklem 3.3'te verilmiştir:

$$h(f_1, f_2, t_2 - t_1) = (t_2 - t_1) \cdot 2^{16} + f_1 \cdot 2^8 + f_2 \quad (3.3)$$

3.2 Ses Parmak İzlerinin Karşılaştırılması

Üretilen ses parmak izleri bu adımda karşılaştırılır ve sonuç olarak benzerlik kriterlerini karşılayan ses dosyasının tespiti gerçekleştirilir. Geleneksel karşılaştırma yöntemlerinde, bu işlem genellikle Şekil 3.8 de gösterildiği gibi, sorgusu yapılan dosyadaki tüm seçkin noktaları, tüm seçkin nokta veri tabanı boyunca, iki boyutlu matrisler şeklinde, ses dosyalarının birim uzunluklar halinde birbirleri üzerinde kaydırılması yoluyla gerçekleştirilmektedir. Ardından, maksimum eşleşme noktası sayısına göre, doğru eşleşme sonucu belirlenmeye çalışılır. Ancak bu tür yöntemler pek tercih edilmezler. Çünkü bu tür yöntemler çok fazla işlem gücü ve zaman gerektirir. Bu yöntemde veri tabanı boyutları arttıkça bu konudaki problemlerde artacaktır.



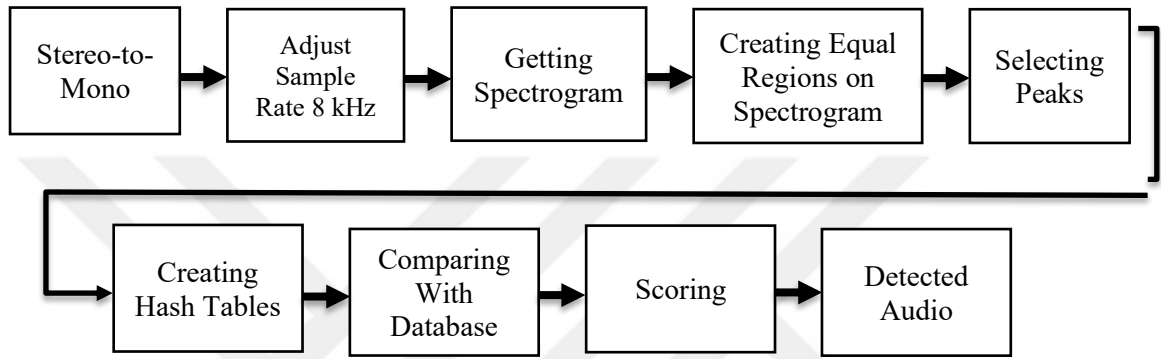
Şekil 3.8 : Geleneksel karşılaştırma tekniği [142].

[3] nolu kaynakta önerildiği gibi, bu iki boyutlu matrislerden üretilecek hash değerleri ile yapılan sorgulama işlemi hem zaman hem de işlem miktarları açısından köklü bir fark yaratmıştır. Bu teknik ile mevcut verilerin entropisi artırılarak güçlü bir çözüm oluşturulmuştur. Aynı prosedürle üretilen veri tabloları (Şekil 3.7-(d)) yardımıyla karşılaştırma işlemi oldukça basit ve hızlı bir şekilde gerçekleştirilir. Bu noktada, Şekil 3.7-(d) de bir örneği verilen, sorgulanan ses dosyasına ait veri tablosundaki (sağda) hash değerleri üstten başlanarak veri tabanındaki (solda) hash değerleriyle karşılaştırılıp eşitlik aranır. Bulunan ilk eşitlikte, hem veri tabanındaki hash değerine karşılık gelen zaman değeri hem de sorgulanan ses dosyasına ait zaman değerleri kaydedilir. Daha sonra bu kaydedilen zaman değerleri ile yeni bir zaman sabiti değeri hesaplanarak iki sütundan oluşan yeni bir tablonun ilk sütununa kaydedilir. İkinci sütunda ise hash değer eşitliği sağlanan veri tabanındaki dosyanın kimlik bilgisi kaydedilir. Bu şekilde tüm karşılaştırma işlemleri tamamlanınca, iki sütundan oluşan bu tablo ile skoreleme işlemine geçilir. Bu noktada son işlem olarak veri tabanındaki her bir dosyanın toplamış olduğu puanlar neticesinde, daha önceden belirlenmiş benzerlik kriterini sağlayan dosya, sorgusu yapılan dosya olarak belirlenir.

4. ARAŞTIRMA BULGULARI VE TARTIŞMALAR

4.1 Eşik Değerli Spektral Nokta Seçiminin Gürültü Gürbüzlüğü

Çalışmaya ait yapılan bu testte, belirli bir eşik değeri ile seçilen spektral noktaların gürültü karşısında gürbüzlüğü incelenmiştir. Bunun için metodoloji bölümünde anlatılan adımlar uygulanmış ve test sonuçları kaydedilmiştir. Uygulamaya ait izlenen adımların akış diyagramı Şekil 4.1 de verilmiştir.



Şekil 4.1 : Uygulamaya ait akış diyagramı.

Çizelge 4.1, rastgele seçilen on adet çoklu ortam ses dosyasından oluşturulan bir veri tabanının özelliklerini göstermektedir. Tüm çoklu ortam ses dosyaları 48 kHz örnekleme hızında, çift kanallı (stereo) yapıda ve MP3 formatındadır.

Çizelge 4.1 : Veri tabanındaki çoklu ortam ses dosyalarına ait özellikler.

File ID	Song Name	Sample Size	Codec	Sample Rate (Fs)
1	Queen-Bohemian Rhapsody	15853568x2	MP3	48 kHz
2	Bob Dylan-Knockin' On Heaven's Door	7268344x2	MP3	48 kHz
3	Salif Keita-Tomorrow	8671572x2	MP3	48 kHz
4	The Beatles-Let It Be	11665592x2	MP3	48 kHz
5	Nirvana-Smells Like Teen Spirit	13361074x2	MP3	48 kHz
6	Coldplay-Paradise	12482803x2	MP3	48 kHz
7	U2 – One	13372220x2	MP3	48 kHz
8	Coldplay-Viva La Vida	11621250x2	MP3	48 kHz
9	Coldplay-Yellow	13093023x2	MP3	48 kHz
10	The Beatles-Yesterday	5548956x2	MP3	48 kHz

Çalışmanın bu adımında, Şekil 4.1 de paylaşılan adımlar uygulanmıştır. Öncelikle metodoloji bölümündeki uygulama adımları ile yukarıdaki veri tabanından ses parmak izi veri tabanı oluşturulmuştur. Bunun için her bir dosya aynı işlemlere tâbi tutulmuş ve elde edilen kimlikler kaydedilmiştir. Bu işlemler tamamlandıktan sonra, sorgusu yapılacak çoklu ortam ses verisine ait kimlik, yani parmak izi, yine birebir aynı prosedür ile hesaplanarak, ses parmak izi veri tabanı ile karşılaştırılmıştır.

Çalışmanın bu adımında eşleştirme testleri, veri tabanından rastgele seçilen bir çoklu ortam ses dosyasının (4-The Beatles - Let It Be) rastgele seçilen 10 saniyelik kısmına, farklı seviyelerde beyaz Gauss gürültüsü (AWGN-Additive White Gaussian Noise) eklendikten sonra gerçekleştirilmiştir. AWGN, sinyallere dengeli bir şekilde uygulanabilen, en yaygın kullanılan gürültü türlerinden birisidir. Ortalama değeri sıfır olan rastgele bir Gauss gürültüsü, Gauss dağılımının DC bileşenine sahip olmadığı anlamına gelir. Gauss sürecinin standart sapması, gürültünün RMS (Root Mean Square) genliği ile ilgilidir [143]. Çeşitli gürültü oranları (SNR-Sinyal gürültü oranı) altında oluşturulan sorgusu yapılacak çoklu ortam ses dosyasının sorgulanan kısmına ait elde edilen hash değer bilgileri Çizelge 4.2 de verilmiştir.

Çizelge 4.2 : Çeşitli gürültü oranlarına karşılık oluşan hash değeri veri sayıları.

SNR (dB)	Hash Table Size
-7	5499x2
-5	5007x2
-3	4692x2
0	2760x2
2	3882x2
6	3264x2
8	3000x2

Bu tablodaki hash değer sayılarının, aynı dosyanın aynı 10 saniyelik kısmına, aynı prosedür uygulanmasına rağmen farklı çıkmasının sebebi, spektral noktalar seçilirken herhangi bir sayı sınırlandırması uygulanmadığı için farklı gürültü seviyelerinde farklı değerler elde edilmesinden dolayıdır. Bu noktalar seçilirken, spektrogram önceden belirlenen bir çerçeve boyutuna göre eş alanlara ayrılır. Daha sonra bu alanların her biri kendi içerisinde bir değerlendirmeye tâbi tutulur. Bu adımda bu değerlendirme işlemi şu şekilde gerçekleştirilmiştir: Öncelikle her alan içerisindeki bileşenlerden genlik değeri bakımından en büyük bileşenin tespiti gerçekleştirilir. Ardından belirlenen bir yüzde eşik

değer ile diğer bileşenler bir değerlendirme işlemine tâbi tutulur. Bu test esnasında bu yüzde değer 90 olarak belirlenmiş ve testler bu değere göre gerçekleştirilmiştir. Yani alan içerisinde kalan bileşenlerin genlik değerleri, en büyük genlik değere sahip bileşenin genlik değerinin %90'ı ve üzerinde olması durumunda, bu bileşenler spektral tepe olarak değerlendirilmek üzere kaydedilir. Bu işlem tüm spektrogram verisine uygulanarak seçkin noktalar haritası (constellation map) oluşturulur. Daha önce de bahsedildiği gibi bu haritaya ait bileşenlerin genlik değeri bilgileri, bu adımdan sonra kullanılmamaktadır ve bu sebeple bu adımdan itibaren kayıt altında tutulmazlar. Bu noktalara ait zaman ve frekans değerleri ile kalan işlemler sürdürülür.

Tabloda dikkat edilecek bir diğer nokta ise gürültü seviyelerinin artması ile oluşturulan hash değer sayıları arasındaki paralelliktir. Bu durumun temel sebebi, artan gürültü seviyeleri ile ses verisine ait ayırt edici bileşenlerin, yani yüksek genlikli noktaların baskılanması sebebiyle, bileşenlerin birbirine yakınsamasından dolayıdır. Bu durum belirlenen eşik değere daha fazla sayıda bileşenin yaklaşmasına sebep olarak daha fazla sayıda hash değer verisi üretilmesine, yani kimlik bilgisinin uzamasına sebep olmuştur.

Elde edilen karşılaştırma sonuçları Çizelge 4.3 te sunulmuştur. Bu tablodan görüldüğü üzere uygulanan algoritma oldukça tatmin edici sonuçlar vermiştir. Tüm farklı gürültü değerleri için, doğru dosyanın tespiti sorunsuz bir şekilde gerçekleştirilmiştir. Beklendiği gibi gürültü seviyesindeki artışa bağlı olarak eşleme sayılarında normal seviyede düşüşler olmuştur, ancak bu durum sorgulanan ses verilerinin doğru tespitini olumsuz bir şekilde etkilememiştir.

Çizelge 4.3 : Gürültü seviyelerine karşı elde edilen eşleşme sayıları.

#	File ID									
SNR (dB)	1	2	3	4	5	6	7	8	9	10
-7	37	28	26	109	37	40	33	28	39	39
-5	45	28	21	120	44	55	38	30	41	38
-3	46	31	24	173	46	44	53	27	50	46
0	35	43	21	513	34	62	70	68	52	44
2	51	34	24	234	34	54	55	26	51	41
6	45	38	22	323	38	50	70	34	53	40
8	49	36	23	265	32	56	42	30	38	44

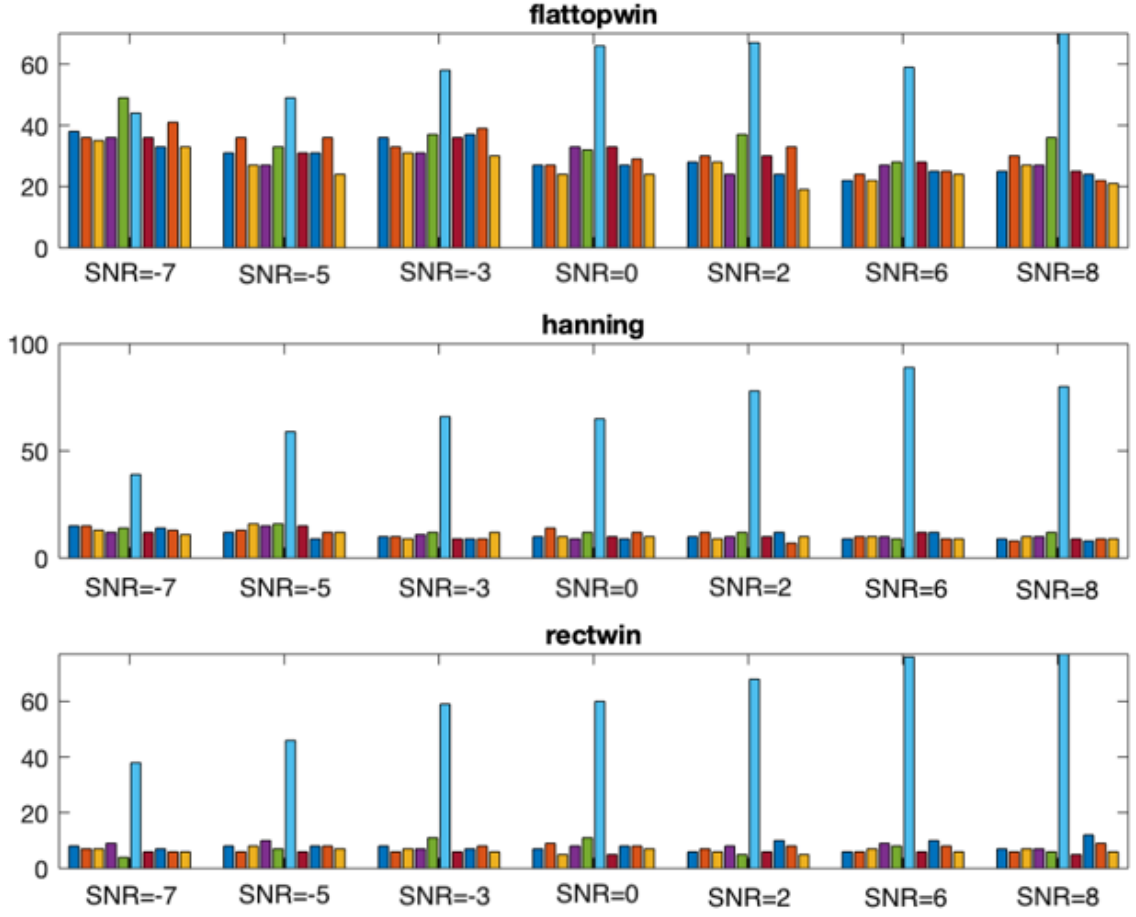
Çizelgeden görüleceği üzere, bu test ile çalışmada önerilen algoritmanın tatmin edici bir şekilde çalıştığı ve beklendiği gibi yüksek gürültü oranlarında bile yüksek oranda eşleşme sağladığı sonucuna varılmıştır. Eşleştirme adımı kullanılarak hash fonksiyonları, tüm

gürültü seviyeleri için iyi sonuçlar vermiştir. Gürültü seviyesindeki artış, algoritmanın doğru eşleşmeyi bulma yeteneğini çok fazla etkilememiştir.

4.2 Farklı Pencere Fonksiyonlarının Algoritmanın Performansı Üzerindeki Etkileri

Çalışmanın bu adımında, çoklu ortam ses dosyalarının tanımlanması için oluşturulan ses parmak izi temelli algoritma içerisinde, spektrogram verileri elde edilirken kullanılan pencere fonksiyonu parametresinin, farklı gürültü seviyeleri dikkate alınarak algoritma üzerindeki etkileri araştırılmıştır.

Bir üstteki uygulamada izlenen metodoloji adımları (Şekil 4.1) birebir aynı şekilde uygulanmıştır. Çalışmanın bu adımı için, Şekil 4.1 de uygulanan adımlar içerisinde, spektrogramın elde edilmesi basamağında kullanılan pencere fonksiyonu parametresi, üç farklı pencere fonksiyonu için ayrı ayrı hesaplamalara dahil edilmiş ve performansa olan etkileri araştırılmıştır. Bu amaçla, seçilen pencere fonksiyonunun işlevinin farklı seviyelerdeki AWGN gürültü değerlerine karşın eşleşme oranları dikkate alınarak karşılaştırılmıştır. Bu pencere fonksiyonları Bölüm 2.4'te özellikleri paylaşılan Flattop, Hanning ve rectangular pencere fonksiyonlarıdır. Farklı gürültü seviyeleri altında üç pencere fonksiyonu için elde edilen uygulama sonuçları Şekil 4.2 de gösterilmektedir.



Şekil 4.2 : Üç farklı pencere fonksiyonu için farklı gürültü seviyelerine karşı elde edilen eşleşme sayıları.

Şekil 4.2, 10 adet çoklu ortam ses dosyasından oluşan bir veri tabanından rastgele seçilen bir ses dosyanın 10 saniyelik herhangi bir bölümü ile gerçekleştirilen sorgulama işlemlerine ait sonuçları göstermektedir. Bu adımda, uygulanan üç farklı tip pencere fonksiyonunun algoritma üzerindeki performans sonuçları ile bu pencere fonksiyonlarının teorideki spektral özellikleri arasında bir karşılaştırma gerçekleştirilmiştir.

Veri tabanındaki 6 nolu çoklu ortam ses dosyasının rastgele seçilmiş 10 saniyelik bir kısmının, 7 farklı seviyedeki gürültü oranı ile yeniden oluşturularak, tüm veri tabanı boyunca bu 10 saniyelik kısmın ses parmak izi ile sorgusu gerçekleştirilmiştir.

Görüldüğü üzere Hanning penceresi ile yapılan simülasyon sonuçları doğru eşleştirme sayısı açısından daha başarılı ve tatmin edici sonuçlar oluşturmuştur. Çünkü, Bölüm 2.4'te de belirtildiği üzere bu pencere fonksiyonu, diğer iki pencere fonksiyonuna göre nispeten daha dar ana loba ve daha düşük yan lob sönümleme oranına sahip olması

nedeniyle mümkün olan en düşük spektral sızıntıyı gerçekleştirerek en iyi sonuçları oluşturmuştur.

Flatlop pencere fonksiyonu, bu uygulamadaki en kötü sonuçların elde edildiği pencere fonksiyonu olmuştur. Bu pencere fonksiyonu, düşük seviyeli geçiş bandı dalgalanması (<0.01 dB) nedeniyle genellikle kalibrasyon amacıyla kullanıldığından bu tür uygulamalarda tatmin edici sonuçlar oluşturamamaktadır. Bu pencerenin bant genişliği, Hanning penceresine kıyasla yaklaşık 2,5 kat daha geniştir. Bu nedenle daha az seçicilik sağlar. Bu durum yapılan testlerde açıkça gözlemlenmiştir.

Dikdörtgen (rectangular) pencere fonksiyonu, Flatlop pencere fonksiyonuna göre daha iyi sonuç verirken, gerçek eşleşme sayısı göz önüne alındığında Hanning penceresine göre daha kötü sonuçlar vermektedir. Normalize edilmiş doğruluk oranı açısından bakıldığında zaman, dikdörtgen pencere gürültü faktörleride dikkate alındığında iyi ve kararlı sonuçlar vermiştir.

Dikdörtgen pencere ile karşılaştırıldığında, Hanning penceresinin gürültü bant genişliği %50 daha yüksektir. Bu nedenle, geniş bantlı rastgele sinyaller için güç spektrumu değerleri, dikdörtgen bir pencere yerine Hanning kullanılarak analiz edildiğinde 1,5 kat daha yüksek olacaktır. Bu nedenle Hanning penceresi, seçicilik, geçiş bandı dalgalanması ve görünür sızıntı açısından dikdörtgen pencereden daha iyi performans gösterir [65]. Bu durum, eşleşme sayılarındaki farklılık ile test sonuçlarında ortaya çıkmıştır.

Dijital sinyal işleme alanında kullanılan pencere fonksiyonlarının kullanımının sağladığı avantajlar yadsınamaz. Spektral özelliklerin incelenmesi sırasında, tüm uygulamalarda en iyisi olabilecek bir pencere fonksiyonu tanımlamak mümkün değildir.

Çalışmanın bu adımı, ses parmak izi tabanlı ses arama için geliştirilen algoritmada kullanılan üç farklı pencere fonksiyonunun etkileri, farklı gürültü seviyelerinin etkisi altında karşılaştırmalı olarak incelenmiştir. Elde edilen sonuçlara göre en yüksek eşleşme skoru, sahip olduğu yan lob yuvarlama oranı ve dar ana lob avantajları nedeniyle Hanning penceresinden elde edilmiştir. Öte yandan, eşleşmelerin orantısal doğruluğu incelendiğinde, dikdörtgen pencere fonksiyonunun eşleşme sayılarının normalize edilmesi durumunda Hanning penceresine en yakın sonuçları verdiği gözlemlenmiştir. Özellikle gürültü seviyeleri dikkate alınır, dikdörtgen pencere fonksiyonunun normalleştirilmiş doğruluk oranı daha kararlı sonuçlar oluşturmuştur.

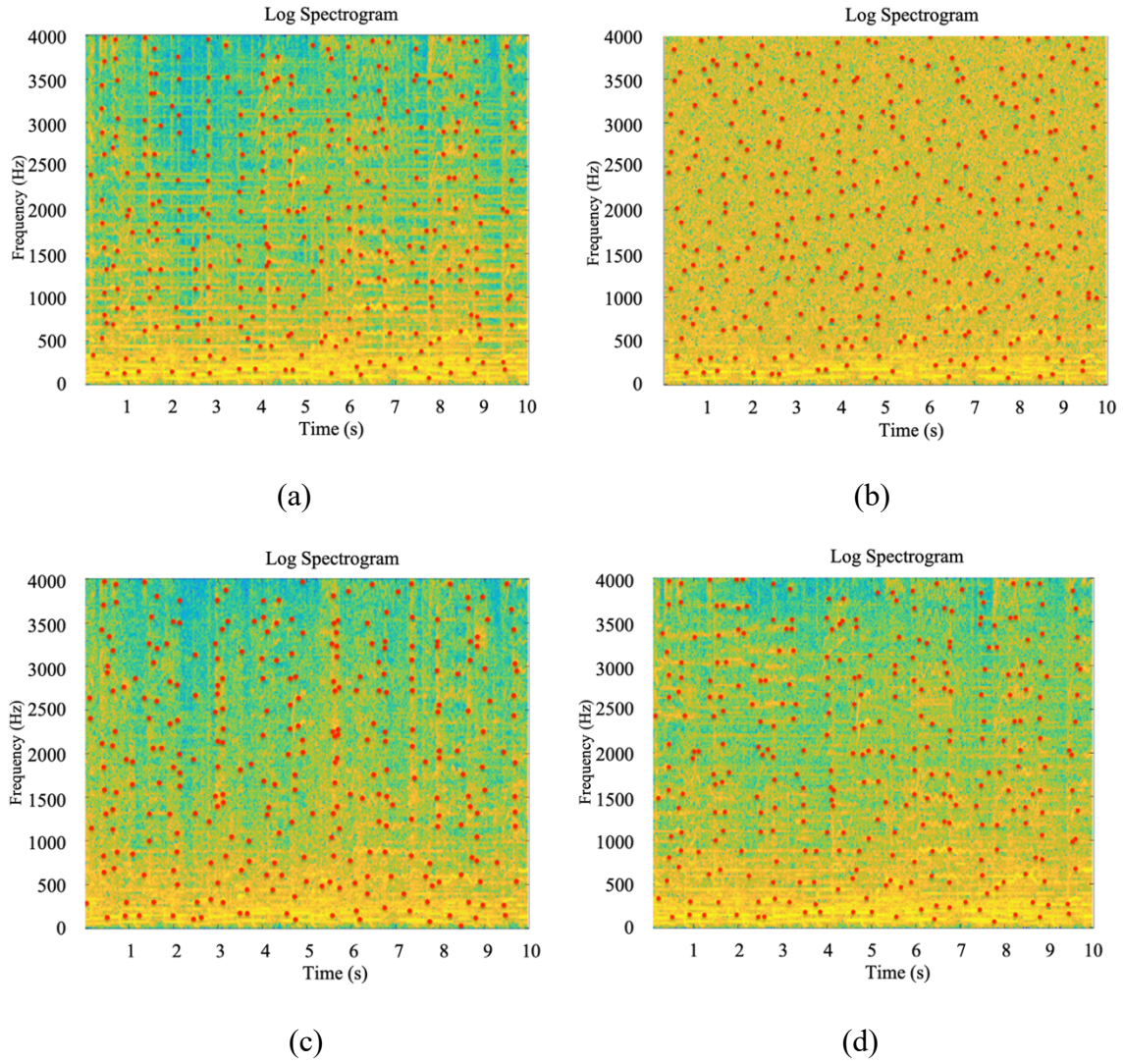
4.3 Bazı Spektrogram Parametrelerinin Etkilerinin Araştırılması

Çalışmanın bu adımında, 50 adet çoklu ortam ses dosyasından oluşan bir veri tabanından rastgele seçilen 10 adet dosyanın her birinin rastgele seçilen 10 saniyelik bölümlerine çeşitli türlerde ve seviyelerde gürültüler (AWGN, restoran ambiyansı, trafik ambiyansı) eklenmiştir. Bu işlem, üç gürültü türü için SNR değerlerinin 3 dB'lik adımlarla -6 dB ile 12 dB arasında ayarlanmasıyla gerçekleştirilmiştir. Veri tabanı, R&B, alternative rock, blues, rock ve dance gibi 5 tür müzik dosyası içermekte ve veri tabanında her müzik türünden on adet şarkı vardır. Test prosedüründe kullanılan ses dosyalarının tür bilgileri Çizelge 4.4 te verilmiştir.

Çizelge 4.4 : Test edilen ses dosyalarının müzik türleri.

Music Genre	Song ID in Database
R&B	4 and 9
Alternative Rock	15 and 18
Blues	21 and 26
Rock	33 and 37
Dance	40 and 42

Çalışmanın bu bölümünde, önceki iki uygulama adımından farklı olarak, spektrogram parametrelerinin etkilerinin en iyi şekilde incelenebilmesi amacıyla, test şartlarının olabildiğince eşit olması adına, tüm dosyalar için saniye başına oluşturulacak spektral nokta sayısı eşit olacak şekilde algoritmada bazı değişiklikler yapılmıştır. Bu durum için belirlenen saniye başına 30 adet spektral nokta seçimi tüm veri tabanı boyunca uygulanmıştır. Kalan tüm adımlar, önceki iki uygulamada olduğu gibi (Şekil 4.1) birebir aynı şekilde uygulanmıştır.



Şekil 4.3 : Veri tabanı içerisinde bulunan bir çoklu ortam ses dosyasının (Song ID:18) 10 saniyelik bir kısmına uygulanan farklı tür ve seviyelerdeki gürültü ekleme işlemi sonucu elde edilen spektrogram çıktıları. Kırmızı noktalar spektral tepeleri temsil etmektedir: (a) Gürültüsüz durum, (b) 3 dB AWGN, (c) 4.8 dB restoran ambiyans gürültüsü, (d) 13.5 dB trafik ambiyans gürültüsü.

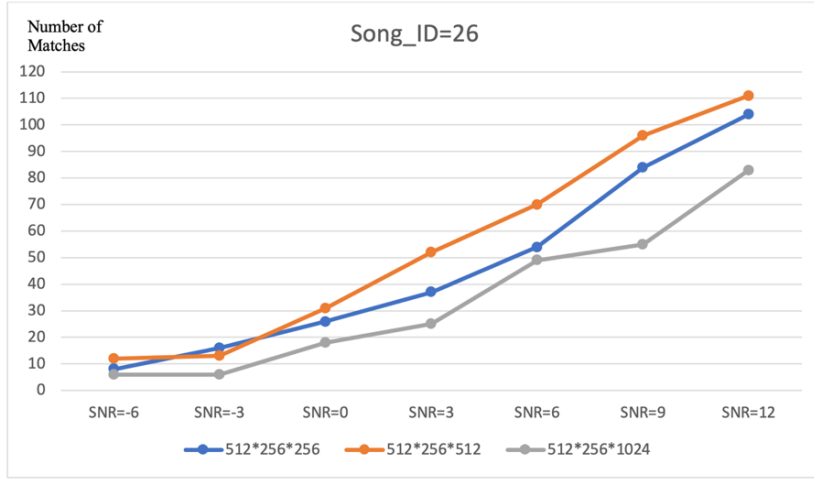
Spektrogram parametrelerinin etkilerini belirlemek için her seferinde pencere boyutu (Window Size), üst üste binme oranı (Overlap) ve FFT sayısı (Number of FFT) parametreleri değiştirilip sonuçlar kaydedilmiştir. Bu parametreler ile 11 farklı kombinasyon oluşturulup testler gerçekleştirilmiştir. Parametre değişim sürecinde her zaman iki parametre sabit tutulurken diğer parametrenin farklı değerleri için testler sürdürülmüştür. Ayrıca bu parametrelerin testleri farklı türde ve eşit seviyede gürültü eklemeleri ile daha detaylı incelenmiştir. Test boyunca eşleşme sayıları ve işlem süreleri kaydedilmiştir.

Yapılan testlerin sonuçları, 10 adet farklı sorgulama işlemine ait eşleşme ve işlem süre değerleri dikkate alınarak daha güvenli bir bakış açısı için bu değerlerin ortalamaları

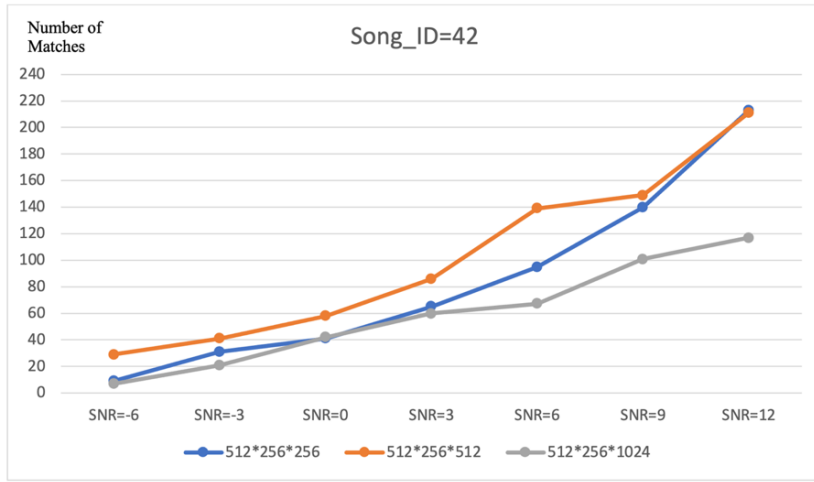
hesaplanarak incelenmiştir. Böylece yapılan bir sorgulama işlemi ile sonuca varmak yerine yapılan tüm sorgulamaların sonuçları değerlendirilmiştir.

Yukarıda bahsedildiği şekilde 10 farklı dosyadan elde edilen hesaplama sonuçlarını paylaşmadan önce, sonuçlar hakkında bir fikir edinebilmek adına iki farklı dosyanın gürültüsüz koşullarda FFT sayısı parametresi için elde edilen sonuçları ayrı ayrı Şekil 4.4 te verilmiştir.

Bu iki grafik incelendiğinde, sonuçların aynı parametre kombinasyonlarında birbirine paralel değerlerde çıktığı açıkça gözlenmektedir. FFT sayısının 512 değerinde seçilmesi iki farklı dosya için hemen hemen tüm gürültü değerleri altında en yüksek eşleşme sayılarının elde edildiği durum olmuştur. Bu konuda bu üç durum göz önüne alındığında, 1024 FFT sayısı değeri en düşük eşleşme sayılarının elde edildiği durum olmuştur.



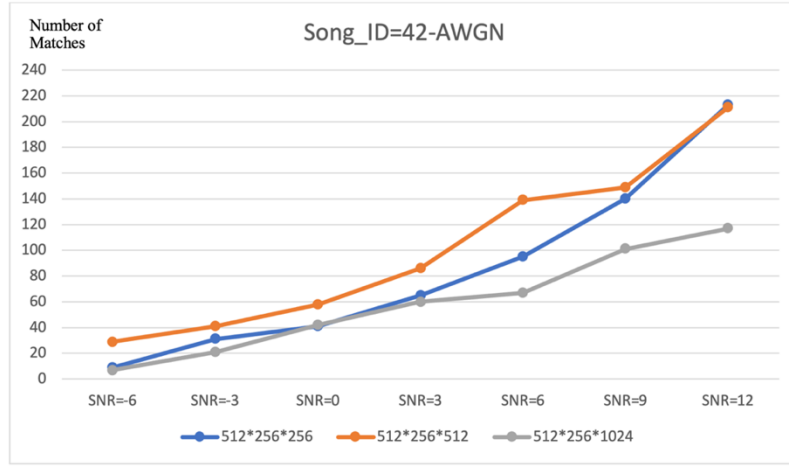
(a)



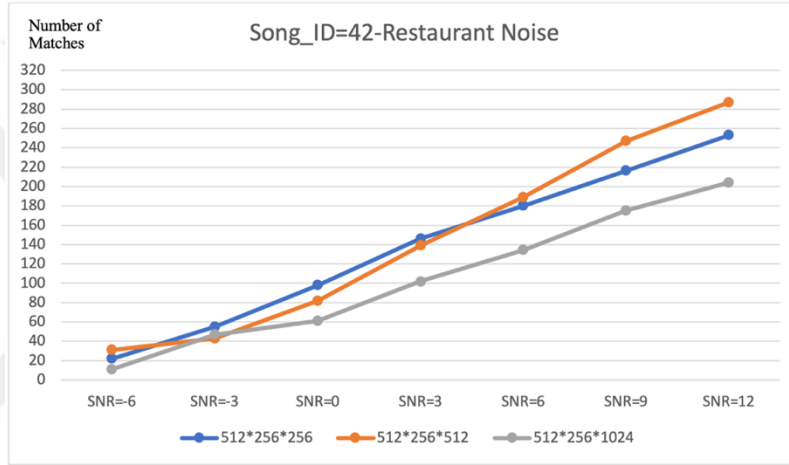
(b)

Şekil 4.4 : Farklı iki ses dosyasına ait AWGN gürültüsü ile gerçekleştirilen FFT sayısı parametresine ait test sonuçları. (a) 26 nolu dosya için sonuçlar, (b) 42 nolu dosya için sonuçlar.

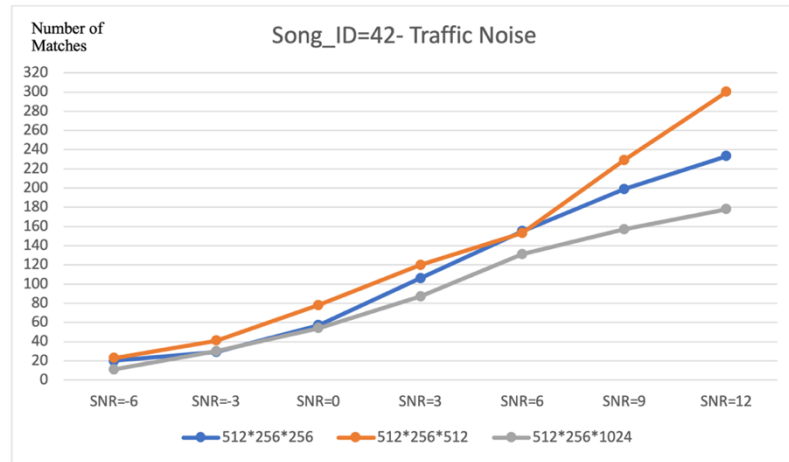
Altta paylaşılan sonuçlar ise veri tabanı içerisinde rastgele seçilmiş bir çoklu ortam ses dosyasına (42 nolu dosya) ait sırasıyla FFT sayısı, üst üste binme oranı ve pencere uzunluğu parametrelerinin eşleşme sayıları üzerindeki etkilerini göstermektedir. Daha sonra paylaşılacak şekiller ise on farklı çoklu ortam ses dosyası için elde edilen bu değerlerin ortalaması hesaplanarak çizdirilmiştir. Böylece hem daha genel sonuçlar elde edilmesi hem de görsel sunum açısından daha anlaşılır ve sade sonuçların paylaşılması hedeflenmiştir.



(a)



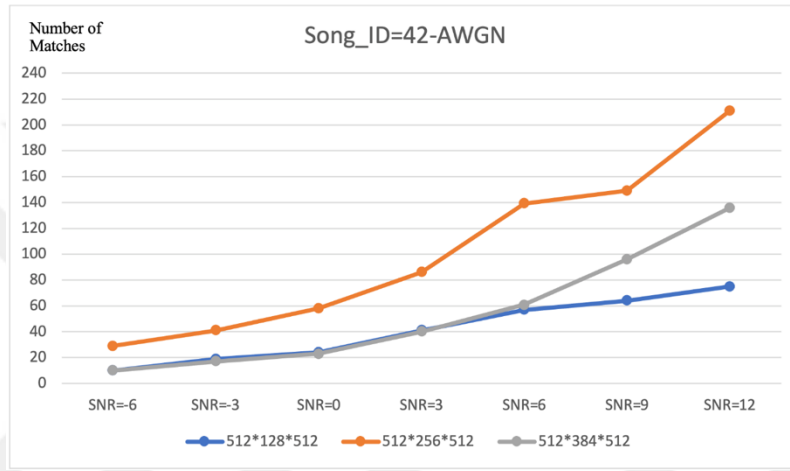
(b)



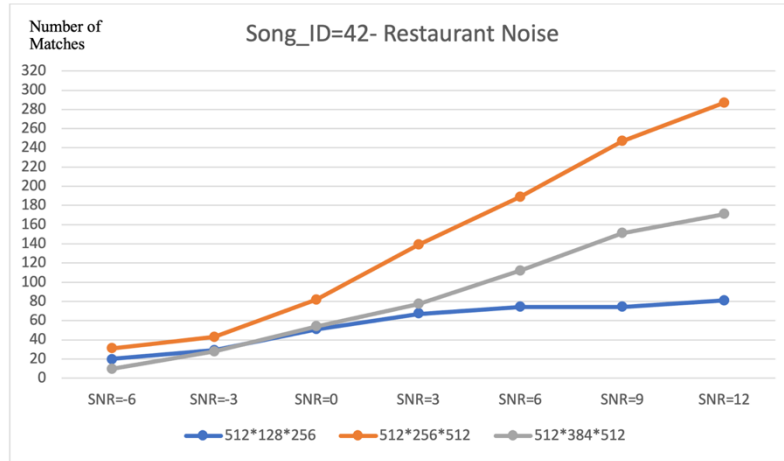
(c)

Şekil 4.5 : Veri tabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç farklı gürültü tipi ile gerçekleştirilen FFT sayısı etkisine ait sonuçlar. (a) AWGN tipi gürültü için sonuçlar, (b) Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar.

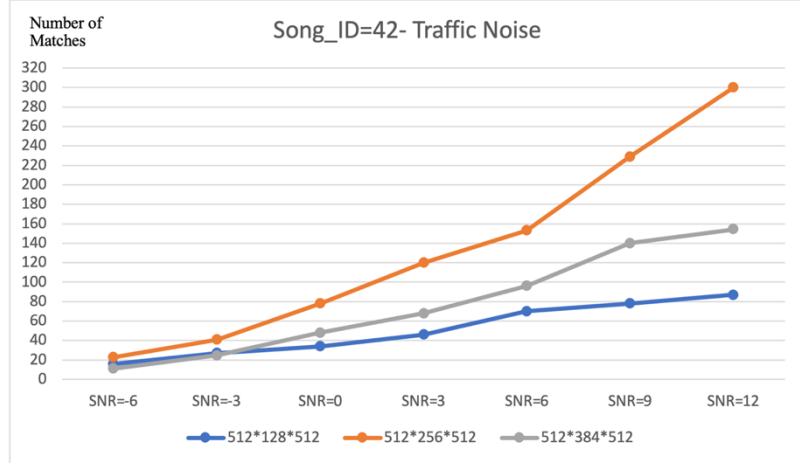
Şekil 4.5 veri tabanı içerisinde rastgele seçilmiş bir çoklu ortam ses doyasına ait üç farklı tipte ve yedi farklı seviyede eklenen gürültüler ile FFT sayısı parametresinin etkisini ölçmeye yönelik yapılan testlerin sonucunu göstermektedir. Üç grafik için 512 seçilen FFT sayısı ortalama eşleşme sayısı açısından en iyi sonuçların elde edildiği değer olmuştur. Restoran gürültüsü ile yapılan testlerde -3 ile 3 dB arasındaki testlerde 256 seçilen FFT sayısı değeri nispeten daha iyi eşleme sayıları sağlamıştır. Ancak geri kalan tüm gürültü seviyeri ve türlerinde 512 değeri daha yüksek eşleşme sayıları sunmuştur.



(a)



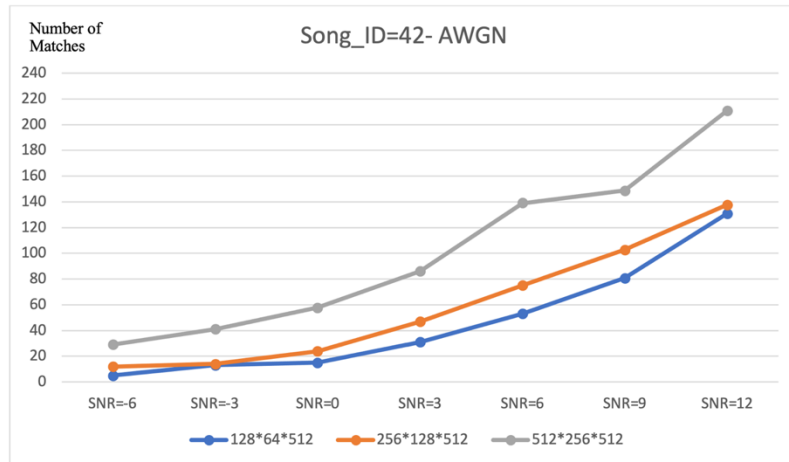
(b)



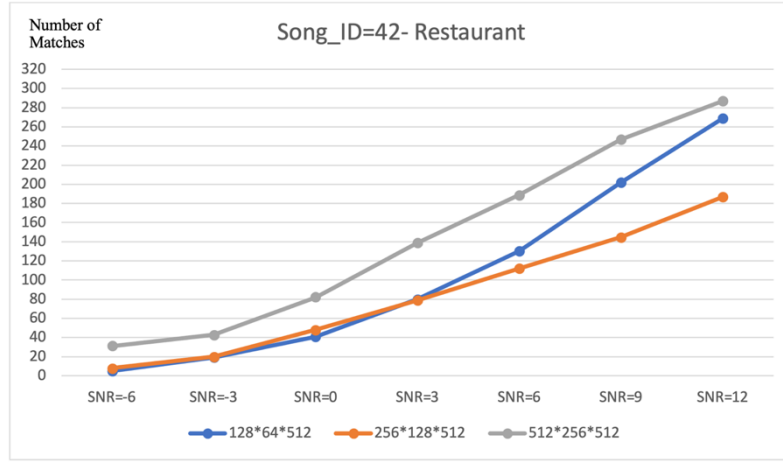
(c)

Şekil 4.6 : Veritabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç farklı gürültü tipi ile gerçekleştirilen üstüste binme oranı etkisine ait sonuçlar. (a) AWGN tipi gürültü için sonuçlar, (b) Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar.

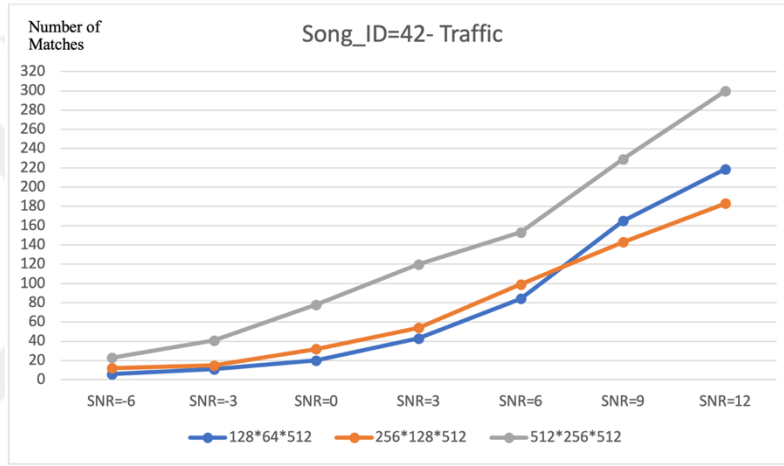
Şekil 4.6 veri tabanı içerisinde rastgele seçilmiş bir çoklu ortam ses dosyasına ait üç farklı tipte ve yedi farklı seviyede eklenen gürültüler ile üstüste binme oranı parametresinin etkisini ölçmeye yönelik yapılan testlerin sonucunu göstermektedir. Grafikler yakından incelenirse, tüm gürültü tiplerinde ve seviyelerinde %50 seçilen üstüste binme oranı eşleşme sayısı bakımından en iyi sonuçların elde edildiği değer olduğu görülecektir. Daha öncede bahsedildiği gibi bu değer %50 ve üzerinde seçilmesi optimum sonuçları vermektedir. Çünkü bu değer altında seçim yapıldığı zaman, bu değerden uzaklaştıkça sinyale ait nitelikli noktalarda azalma meydana gelmektedir. Bu durum bu dosya üzerinde yapılan bu testlerde gözlemlenmiştir.



(a)



(b)



(c)

Şekil 4.7 : Veritabanında bulunan 42 nolu çoklu ortam ses dosyasına ait üç farklı gürültü tipi ile gerçekleştirilen pencere uzunluğu etkisine ait sonuçlar. (a) AWGN tipi gürültü için sonuçlar, (b) Restoran ambiyans gürültüsü için sonuçlar, (c) Trafik ambiyans gürültüsü için sonuçlar.

Şekil 4.7 veri tabanı içerisinde rastgele seçilmiş bir çoklu ortam ses dosyasına ait üç farklı tipte ve yedi farklı seviyede eklenen gürültüler ile pencere uzunluğu parametresinin etkisini ölçmeye yönelik yapılan testlerin sonuçlarını göstermektedir. Grafiklere yakından bakacak olursak, en iyi eşleşme sayısı değerlerine 512 örnek uzunlukta seçilen pencereler ile ulaşıldığı açıkça görülecektir. Bu tür ses dosyalarının 64 ms uzunlukta parçaları, Fourier dönüşüm tekniği için doğru sonuçların elde edilmesi için gerekli olan sinyalin durağan yapıda olması şartını sağlayan değerler aralığında olduğunu göstermektedir. Grafiklerde bu değer en kararlı sonuçların elde edildiği değer olmuştur.

Şekil 4.5, 4.6 ve 4.7 için, her üç şekilde de 512;%50;512 parametre kombinasyonu genel olarak en iyi sonuçların elde edildiği parametre kombinasyonu olduğunu söyleyebiliriz. Bu durum gürültü seviyelerindeki azalma ile daha çok ön plana çıkmıştır.

Veri tabanından seçilen 10 farklı ses dosyanın, 10 saniyelik kısımlarının gürültüsüz ve 3 farklı türde 3 dB gürültü eklenmesi ile oluşturulan formlarına ait, 11 farklı spektrogram parametre kombinasyonu ile yapılan test sonuçları ise Çizelge 4.5 de verilmiştir.

Çizelge 4.5 : Farklı spektrogram parametrelerine göre 10 farklı ses dosyası için 10 saniyelik bölümlerine ait farklı gürültü tipleri ile yapılan testlere ait eşleşme sayılarının ve tanıma sürelerinin ortalamaları.

#	Variable Combinations of Spectrogram Parameters			Noiseless		AWGN SNR: 3 dB		Restaurant Noise SNR: 3 dB		Traffic Noise SNR: 3 dB	
	Window Size	Overlap	Number of FFT	Matching Averages	Average Times (s)	Matching Averages	Average Times (s)	Matching Averages	Average Times (s)	Matching Averages	Average Times (s)
1	128	50%	512	636	5.565	32.4	4.788	107.9	5.389	55.2	5.264
2	256	50%	256	248.7	7.877	42.3	6.233	86.2	7.081	67.1	6.784
3	256	50%	512	271.5	4.698	42.6	4.163	84.4	4.156	69.6	4.164
4	256	75%	256	339.3	5.083	33.3	4.616	86.6	4.694	51.7	4.684
5	256	75%	512	225	4.956	27	4.384	61.4	4.424	47.2	4.479
6	512	25%	512	73	5.752	30.4	5.044	33.3	5.468	32	5.283
7	512	50%	256	337.3	9.924	84.3	7.627	127.5	8.811	98.8	8.26
8	512	50%	512	351.4	5.778	69.8	5.112	113.8	5.304	106.2	5.13
9	512	50%	1024	304.3	4.392	47.8	4.179	91.4	4.253	81.3	4.185
10	512	75%	256	233.9	6.63	41.7	5.778	80.8	6.15	66.1	6.044
11	512	75%	512	256.9	4.385	43.3	4.392	87	4.221	69	4.104

Tablodan görüldüğü gibi, ortalama eşleşme sayıları (NM-Number of Matches) ve ortalama tanıma süreleri, spektrogram parametrelerinin kombinasyonuna bağlı olarak değişiklik göstermektedir. Tanıma süreleri, genel olarak kullanılan sisteme göre tatmin edicidir. En kötü ortalama eşleşme sayıları 512;%25;512 kombinasyonu için 6. satırda elde edilirken, gürültüsüz durumda ilk satırda 128;%50;512 kombinasyonu için maksimum ortalama eşleşme sayısı değeri elde edilmiştir. Ancak gürültülü durumlar dikkate alındığı takdirde, bu kombinasyon, ortalama eşleşme sayısındaki nispeten daha fazla azalma nedeniyle, diğer kombinasyonlar arasından en kötü sonucun elde edildiği kombinasyon olmuştur.

4, 7, 8 ve 9 numaralı satırlarda verilen diğer kombinasyonlar için de yüksek ortalama eşleşme sayısı değerleri elde edilmiştir. 4. sırada verilen kombinasyon için gürültülü durumda ortalama eşleşme sayısı değerindeki azalma oldukça yüksektir. 4, 7, 8 ve 9 sıraları

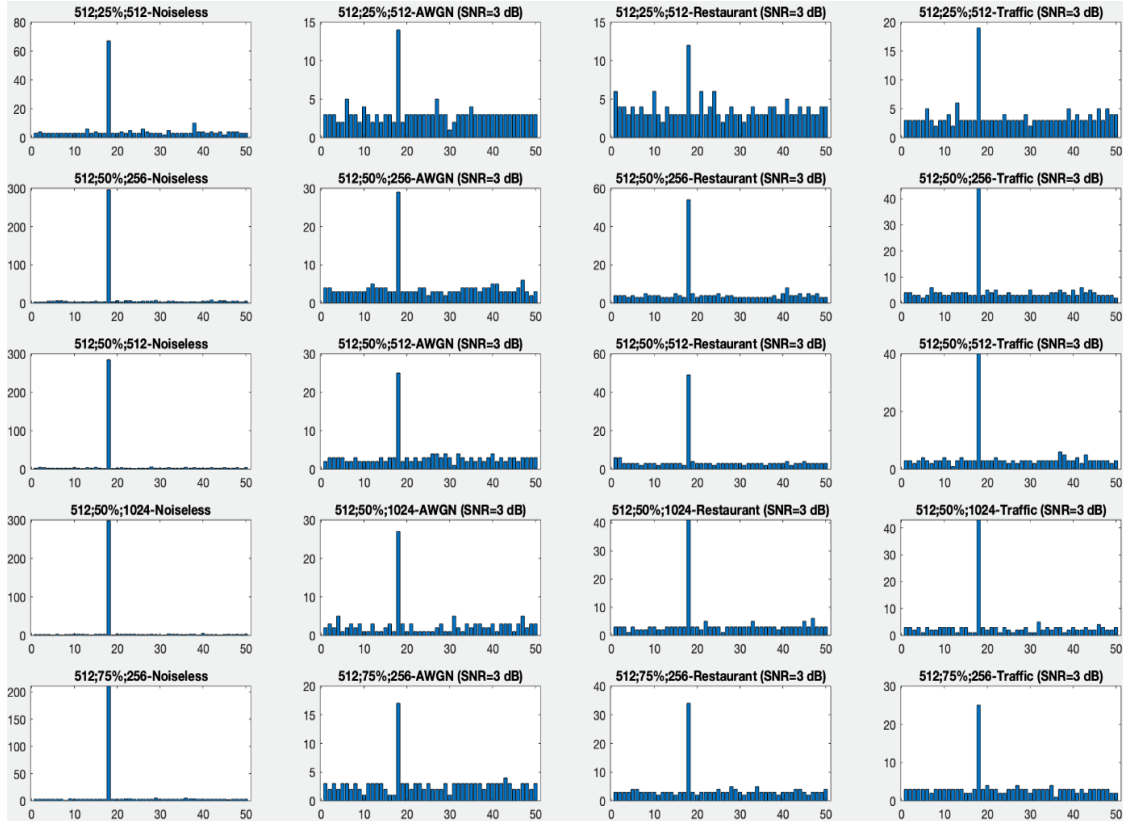
arasında gürültülü durumda ortalama eşleşme sayısı değerinde görece minimum düşüş 7 ve 8 numaralı sıralarda gözlenmiştir.

Dolayısıyla gürültüye karşı gürbüzlük hesaba katıldığında 7 ve 8 numaralı satırlarda verilen spektrogram parametre kombinasyonları diğer kombinasyonlara görece daha iyi sonuç vermektedir. Bu iki satır ortalama tanıma süresi dikkate alınarak karşılaştırıldığında, 8. satırda elde edilen ortalama tanıma zamanı değerlerinin diğerlerine göre daha az olduğu gözlemlenmiştir.

Tablodan da görüldüğü gibi ortalama eşleşme sayıları ve ortalama tanıma süresi değerleri, spektrogram parametrelerinin güvenilirlik ve sağlamlık üzerindeki etkileri hakkında bazı bilgiler vermiştir. Ancak, bu değerler spektrogram parametrelerinin optimum kombinasyonunu belirlemek için tek başına yeterli değildir. Ortalama eşleşme sayısı değerlerinin yanı sıra, farklı gürültü türlerine ve seviyelerine göre bu değerlerde yaşanan düşüş de dikkate alınmalıdır. Bu nedenle gürültü etkisini ve göreceli eşleşme sayısını dikkate alan bir başka kriterin daha iyi bir değerlendirme için kullanılmasının daha faydalı olacağı düşünülmüştür. Bu kriter “normalleştirilmiş göreceli güvenilirlik” (NRR-Normalized Relative Reliability) olarak adlandırılmış ve denklem 4.1 ile verilen formülle hesaplanmıştır:

$$NRR\%=100*(NMD-NMC)/NMD \quad (4.1)$$

NMD, veri tabanında aranan şarkının eşleşme sayısıdır (Number of Matchings of the Song Searched in the Database). NMC ise NMD'ye en yakın puana sahip herhangi bir şarkının eşleşme sayısıdır (Number of Matchings of Any Song Which Has the Closest Score to NMD). Daha iyi bir analiz için bu değer hesaplanmış ve bu hesaplamalara dayalı grafikler yukarıda belirtilen farklı koşullar dikkate alınarak sunulmuştur. Bundan önce veri tabanında 18 nolu şarkı için NM değerlerini veren klasik sunum Şekil 4.8 de gösterilmiştir. Bu şekilde doğru ses dosyasının tespit edilmesinde en önemli gösterge olan güvenilirlik karşılaştırmalı olarak analiz edilmiştir. Gürültüsüz durum, AWGN, restoran ve trafik gürültülerinin 3 dB'lik SNR değerleri dikkate alınarak veri tabanındaki eşleşme sayıları hesaplanmıştır.



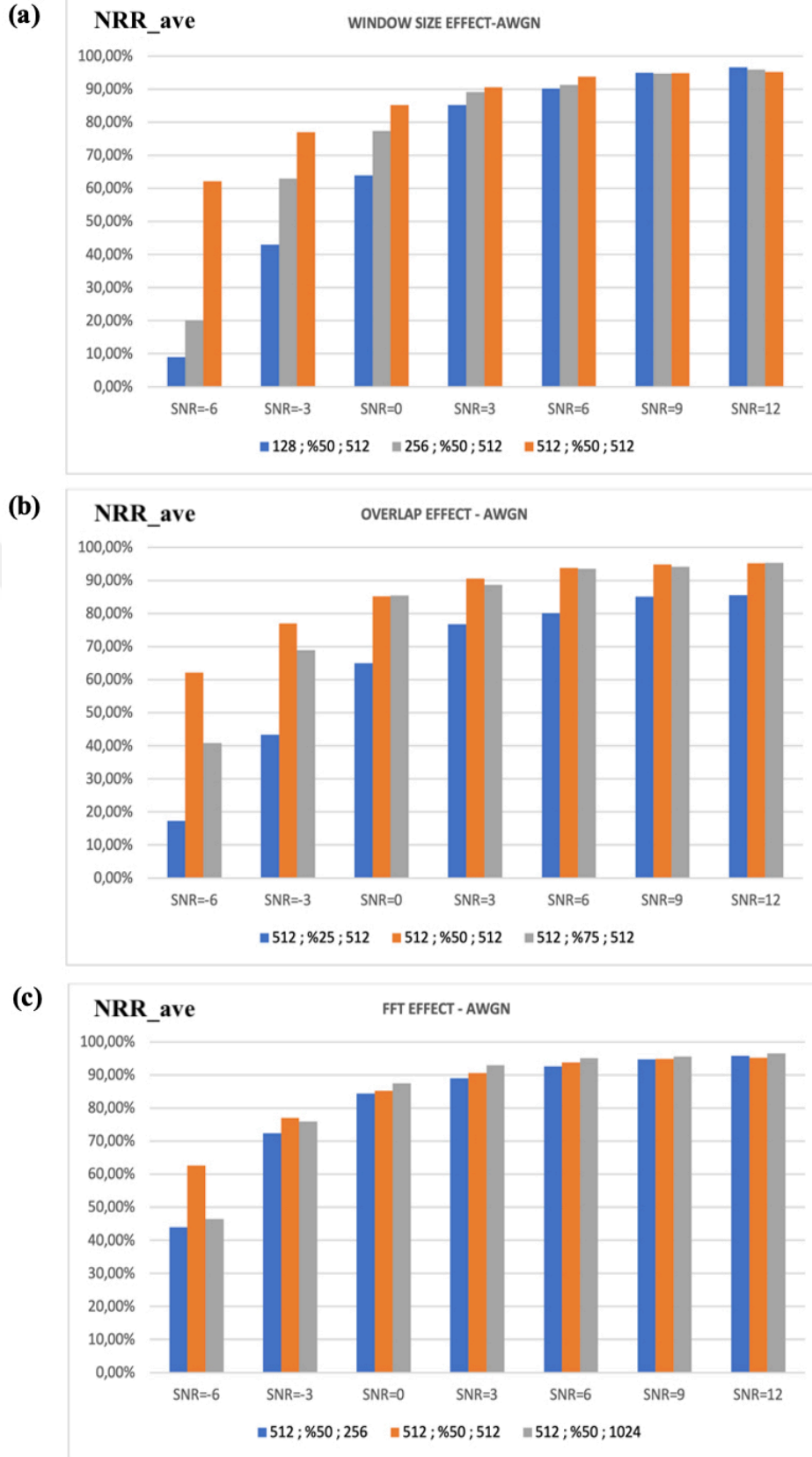
Şekil 4.8 : 18 nolu şarkının rastgele seçilen 10 saniyelik bölümü için oluşan test sonuçları. 5 farklı spektrogram parametre kombinasyonuna göre elde edilen değerler yatay ekseninde dosya numarasını, dikey ekseninde ise eşleşme sayılarını gösterecek şekilde ifade edilmiştir.

Eşleşme sayıları, şekilde görüldüğü gibi beş farklı spektrogram parametresi kombinasyonu için gösterilmiştir ve Çizelge 4.5’de 5,6,7,8,9 satırlarına karşılık gelir. Burada, şarkının rastgele seçilen 10 saniyelik bir bölümü (Song ID=18) dikkate alınmış ve program o şarkıyı farklı spektrogram parametreleri ve gürültü koşulları için veri tabanında bulmada başarı yönünden test edilmiştir. Şekilde görüldüğü üzere program her koşulda istenilen şarkıyı büyük bir doğrulukla bulmuş ve veri tabanında aranan şarkı diğer şarkılar arasında önemli ölçüde ön plana çıkmıştır. Gürültünün artmasıyla birlikte eşleşme sayısının normal olarak azaldığı, gürültünün türüne bağlı olarak azalma miktarının değiştiği görülmektedir.

Eşleşme azalma sayısı dikkate alındığında, en çok ve en az etkili gürültü türleri sırasıyla AWGN ve restoran gürültüsü olarak gözlenmiştir. Her durumda, doğru şarkı, gürültülü ortamlarda bile diğer ses dosyalarıyla karşılaştırıldığında çok yüksek bir eşleşme

oraniyla başarılı bir şekilde tespit edilmiştir. 512;%50;256, 512;%50;512 ve 512;%50;1024 spektrogram parametreleri kombinasyonu için gözlemlenen güvenilirlikler birbirine yakın ve diğer spektrogram parametreleri için elde edilenlerden çok daha iyi oluşmuştur. Bu test başka şarkılar için de yapılmış ve benzer sonuçlar elde edilmiştir. Aranan şarkı ile diğer ses dosyaları arasındaki eşleşme sayılarındaki göreceli farkın büyük olmasından dolayı veri tabanını genişletmenin doğru şarkının tespitinde başarıyı etkilemeyeceği tahmin edilmektedir.

Yukarıda bahsedilen açıklamalar doğrultusunda, spektrogram parametrelerinin farklı gürültü türlerine ve seviyelerine bağlı olarak tanıma sürecine etkisinin daha iyi analiz edilebilmesi için bir sonraki aşamada bir diğer kriter olan NRR dikkate alınmıştır. Bu amaçla Çizelge 4.4 te ele alınan on şarkı için ortalama NRR değerleri hesaplanmış, daha sonra bu değerlerdeki değişimler Şekil 4.9 da değişken spektrogram parametreleri için artan AWGN seviyelerine göre sunulmuştur.



Şekil 4.9 : AWGN gürültüsü için artan SNR seviyelerine karşı (a) pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri için elde edilen ortalama NRR değerlerinin değişimine ait grafikler.

Şekil 4.9 (a) da pencere boyutunun etkisi, değeri 128, 256, 512 olarak değiştirilerek ve üst üste binme oranı ve FFT sayısı parametreleri %50 ve 512 olarak sabit tutularak karşılaştırmalı olarak analiz edilmiştir. Şekil 4.9 (a) da görüldüğü gibi, üç farklı SNR değeri için (6, 9 ve 12 dB) ortalama NRR değerleri birbirine yakın oluşmuş ve pencere boyutundaki değişikliğe bağlı olarak önemli ölçüde farklılık göstermemiştir. Ancak gürültü etkisi arttığında ve SNR değerleri -6, -3 ve 0 dB değerlerine düşürüldüğünde, güvenilirlik ve sağlamlık açısından 512 pencere boyutu değerinin ön plana çıktığı görülmüştür.

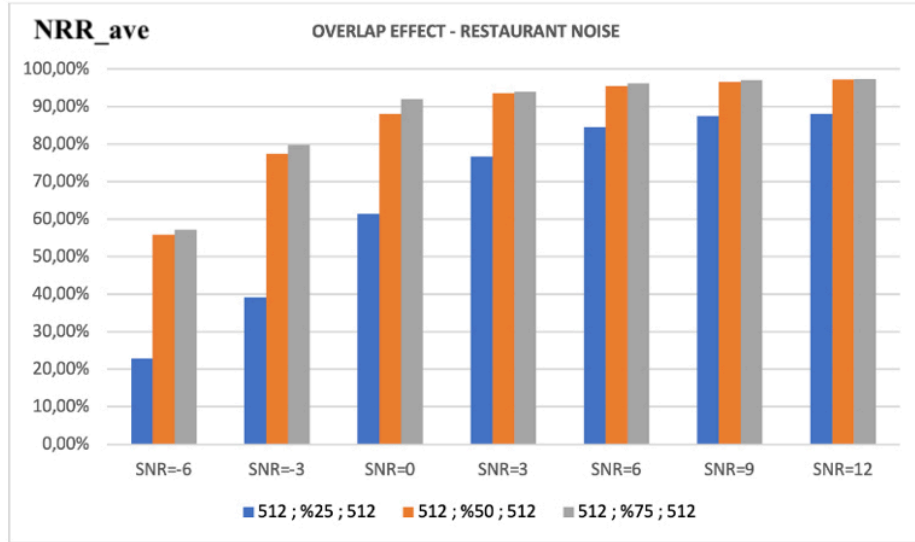
Şekil 4.9 (b) de üst üste binme oranının etkisi, değeri %25, %50, %75 olarak değiştirilerek ve pencere boyutu ve FFT sayısı parametreleri 512 değerinde sabit tutularak karşılaştırmalı olarak analiz edilmiştir. Şekil 4.9 (b) de görüldüğü gibi, SNR değerleri için 0, 3, 6, 9 ve 12 dB, ortalama NRR değerleri %50 ve %75 üst üste binme oranı değerleri için birbirine yakın görünürken, %25 değeri için nispeten daha düşük çıkmıştır. Ancak SNR değerleri -6 ve -3 dB değerlerine düşürüldüğünde %50 lik üst üste binme oranı değerinin ön plana çıktığı görülmüştür.

Şekil 4.9 (c) de FFT sayısının etkisi, değeri 256, 512, 1024 olarak değiştirilerek ve pencere boyutu ve üst üste binme oranı parametreleri 512 ve %50 olarak sabit tutularak karşılaştırmalı olarak analiz edilmiştir. Şekil 4.9 (c) de görüldüğü gibi, -3, 0, 3, 6, 9 ve 12 dB SNR değerleri için ortalama NRR değerleri birbirine yakın oluşmuş ve FFT sayısındaki değişikliğe bağlı olarak anlamlı bir farklılık göstermemiştir. Ancak SNR değeri -6 dB değerine indirildiğinde 512 FFT sayısının ön plana çıktığı görülmüştür. Ayrıca tanıma sürecinde en az etkili spektrogram parametresi olarak FFT sayısı parametresi olduğu gözlemlenmiştir.

(a)



(b)



(c)

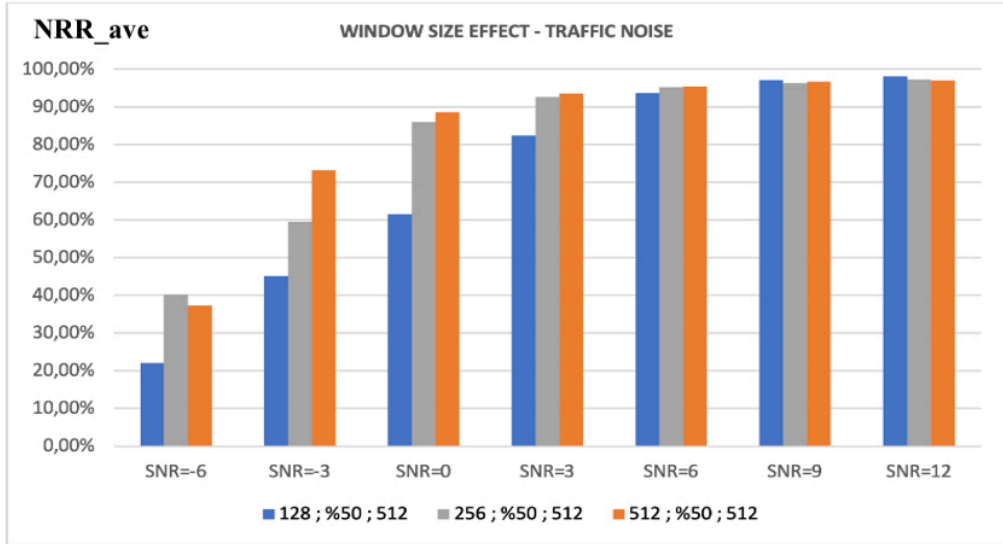


Şekil 4.10 : Restoran gürültüsü için artan SNR seviyelerine karşı (a) pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri için elde edilen ortalama NRR değerlerinin değişimine ait grafikler.

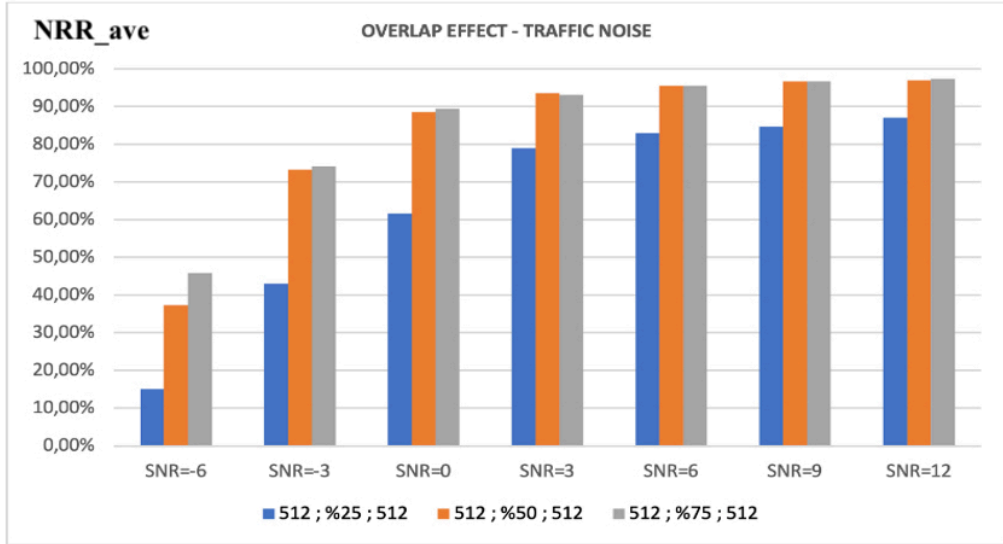
Ses tanıma sürecinde farklı gürültü türlerinin uygun spektrogram parametrelerinin belirlenmesine etkisini gözlemlemek için restoran ve trafik ambiyans gürültülerinin farklı SNR seviyelerinde işleme dahil edilerek benzer analizler yapılmıştır. Şekil 4.10 da, farklı spektrogram parametrelerinin NRR üzerindeki etkisi, restoran gürültüsü için artan SNR değerlerine göre sunulmuştur. Şekil 4.10 (a) da görüldüğü gibi, pencere boyutu yüksek SNR değerleri için düşük gürültü koşullarında anlamlı bir etkiye sahip olmazken, -3 dB ve -6 dB SNR değerleri için, yani yüksek gürültü seviyelerinde, 256 ve 512 değerlerinde bu parametre nispeten öne çıkmaktadır.

Şekil 4.10 (b) de görülen üst üste binme oranı etkisi dikkate alındığında, tüm gürültü seviyelerinde %50 ve %75 oranlarındaki üst üste binme oranı değerleri öne çıkmakta ve en kötü sonuçların %25 üst üste binme oranı değerinde oluştuğu görülmektedir. Şekil 4.10 (c) de FFT sayısı incelenecek olunursa, özellikle düşük SNR seviyeleri için, çok yüksek gürültü koşullarında 256 ve 512 FFT sayıları ön plana çıkmaktadır.

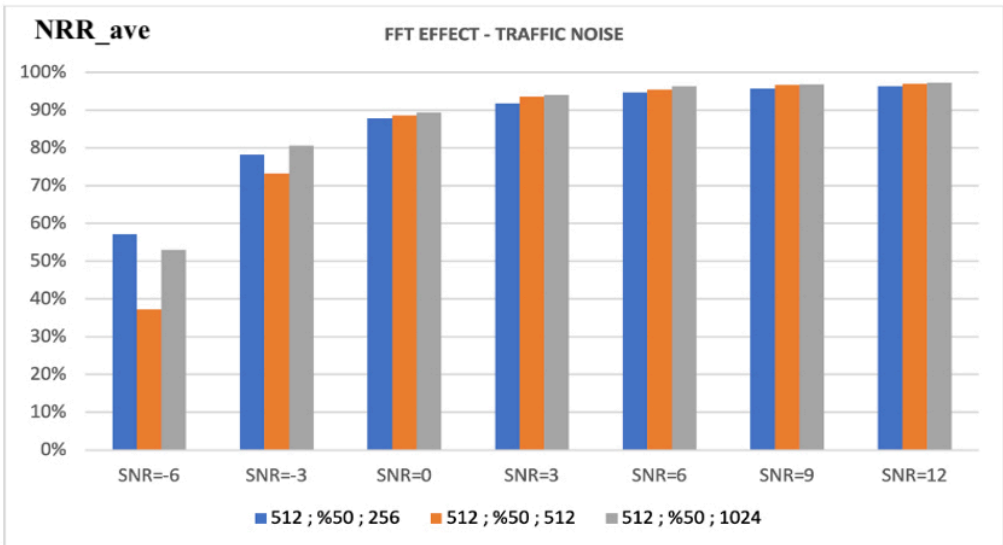
(a)



(b)



(c)



Şekil 4.11 : Trafik gürültüsü için artan SNR seviyelerine karşı (a) pencere boyutu, (b) üst üste binme oranı, (c) FFT sayısı parametrelerindeki değişiklik dikkate alınarak, on farklı şarkının 10 saniyelik bölümleri için elde edilen ortalama NRR değerlerinin değişimine ait grafikler.

Şekil 4.11 de, farklı spektrogram parametrelerinin NRR üzerindeki etkisi, trafik gürültüsü için artan SNR değerlerine göre sunulmuştur. Şekil 4.11 (a) ve 4.11 (b) de görüldüğü gibi, elde edilen sonuçlar AWGN ve restoran gürültüsü için elde edilenlere oldukça benzerdir. Yani 256 ve 512 pencere boyutu değerleri ile %50 ve %75 üst üste binme oranı değerleri, görece yüksek gürültü seviyelerinde ön plana çıkmakta ve düşük gürültü seviyelerinde anlamlı bir etkiye sahip değildir. Şekil 4.11 (c) ye göre FFT sayısının etkisi düşünüldüğünde, özellikle -3 dB ve -6 dB gibi düşük SNR seviyeleri için 256 ve 1024 FFT sayıları ön plana çıkmaktadır. Ayrıca, AWGN ve restoran ambiyans gürültüsü için daha önce elde edilen sonuçlarda gözlemlendiği gibi, FFT sayısındaki değişikliğin normalize edilmiş güvenilirlik üzerinde daha yüksek SNR değerlerinde belirgin bir etkiye sahip olmadığı görülmüştür.

Eşleşme sayısı, normalleştirilmiş bağıl güvenilirlik ve tanıma süresi gibi tüm kriterler birlikte dikkate alınırsa, 512;%50;512, 512;%50;256 ve 256;%75;256 değerlerindeki parametre kombinasyonları görece daha iyi sonuçlar vermiştir. Gürültüsüz durum için Çizelge 4.5 te 10 farklı şarkı için eşleşme sayısını içeren bir karışıklık matrisi sunulmuştur. Matriste spektrogram parametreleri 512;%50;512 olarak alınmıştır. Çizelge 4.5 te görüldüğü gibi, eşleşme sayıları, dolayısıyla doğru şarkıyı bulma olasılığı, belirtilen spektrogram parametreleri kombinasyonu için oldukça yüksektir.

Çizelge 4.6 : Gürültüsüz durumda 512;%50;512 spektrogram parametre kombinasyonuna göre on farklı ses dosyasının 10 saniyelik eşleşme sayılarını gösteren karışıklık matrisi (confusion matrix).

Song ID	4	9	15	18	21	26	33	37	40	42
4	480	3	3	6	4	6	4	4	5	4
9	3	439	2	3	2	3	3	3	3	3
15	3	3	355	3	4	3	3	3	3	5
18	3	3	5	285	4	3	3	3	3	4
21	3	3	3	3	327	7	4	3	8	3
26	3	6	3	3	5	284	5	3	6	2
33	5	3	3	9	6	5	249	3	5	3
37	3	3	3	3	3	3	3	320	3	2
40	3	3	3	3	3	4	3	4	380	2
42	3	3	3	2	3	3	3	3	3	395

Çalışmanın bu adımında, spektrogram parametrelerinin spektral tepeler yöntemine dayalı bir çoklu ortam ses arama algoritmasının güvenilirliği ve sağlamlığı üzerindeki etkileri, farklı gürültü seviyeleri ve türleri dikkate alınarak incelenmiştir. Çalışma için özel

olarak yazılmış programın tanıma süreleri ve eşleşme sayısı yönünden farklı şarkılar test edilerek doğrulanmasının ardından, 10 sn'lik bölümlere pencere boyutu, üst üste binme oranı ve FFT sayısı gibi farklı üçlü spektrogram parametre kombinasyonları, farklı gürültü koşulları altında 50 adet çoklu ortam ses dosyası içeren bir veri tabanından rastgele seçilen 10 farklı çoklu ortam ses dosyasına uygulanmıştır. Bu tür programlar için doğru şarkının kısa sürede tespiti çok önemli olduğundan, değerlendirme sürecinde farklı gürültü koşulları altında eşleşme sayısı, tanıma süreleri ve normalleştirilmiş bağıl güvenilirlik gibi üç farklı kriter dikkate alınmış ve bu kriterler değerlendirilmiştir.

Tüm sonuçlar birlikte analiz edildiğinde, genel olarak spektrogram parametrelerindeki değişikliklerin, 0 dB'den yüksek SNR değerleri için düşük gürültü koşullarında güvenilirliği ve sağlamlığı önemli ölçüde etkilemediği görülmektedir. Ancak gürültü seviyesi kademeli olarak artırıldığında daha düşük SNR değerlerinde 512;%50;512, 512;%50;256 ve 256;%75;256 gibi spektrogram parametre kombinasyonlarının NRR kriterine göre ön plana çıktığı görülmektedir. Uygulama alanına göre bahsedilen kriterlerin önceliğine bağlı olarak bu spektrogram parametre kombinasyonlarından biri seçilebilir. Ancak tanıma süresi, eşleşme sayısı ve gürültüye karşı sağlamlık birlikte dikkate alındığında, 512;%50;512 spektrogram parametre kombinasyonunun en iyi sonuçları verdiği açıkça görülmektedir. Eşleşme sayısı ve güvenilirlik faktörü kriterleri farklı gürültü türlerine göre değişmekle birlikte, bu değişikliğin uygun spektrogram parametre kombinasyonunun belirlenmesinde gözle görülür bir etkisinin olmadığı görülmektedir. Ayrıca müzik türünün eşleşme sayısı üzerindeki etkisi yapılan bu çalışmalarda net bir şekilde gözlemlenememiştir.

5. SONUÇ VE ÖNERİLER

Ses parmak izi, bir ses sinyalinin temsil etmek için sinyalden türetilen bir dijital kimliktir. Ses sinyali işleme alanında sinyallerin doğrulanması, tanınması, etiketlenmesi gibi daha birçok farklı alanda yaygın bir şekilde kullanılmaktadır.

Gerçek hayattaki uygulamalarında, bir ses parmak izinden beklenen en önemli özellikler kompakt bir yapıda olup, çeşitli bozulma durumlarına karşı dayanıklı olması ve gerektirdiği uygulama işlem süresinin mümkün mertebe kısa olması yönündedir.

Bu çalışmada, Shazam'a ait ses parmak izi metodu daha önce bu alanda yapılan çalışmalar baz alınarak geliştirilip, çeşitli türdeki çoklu ortam ses dosyalarına uygulanarak, oluşturulan yöntemin güvenilirliği, kararlılığı ve hızı dikkate alınarak çeşitli testlere tâbi tutulmuştur.

Oluşturulan algoritma öncelikle gürültü gürbüzlüğü açısından test edilmiş ve oldukça tatmin edici sonuçlar elde edilmiştir. Tüm gürültü seviyeleri altında doğru dosyanın tespiti başarılı bir şekilde gerçekleştirilmiştir. Bu uygulamanın ardından algoritmanın işleyişinde oldukça önemli bir yer tutan spektrogram yapısı içerisinde kullanılan pencere fonksiyonlarının algoritma performansı üzerindeki etkileri araştırılmış ve bu konuda en başarılı sonuçlara teoride de beklendiği gibi Hanning penceresi ile ulaşılmıştır.

Bu testlerin ardından, algoritmada önemli bir karar aşaması olan spektrogram parametrelerinin algoritma üzerindeki etkileri uygulanan testler sonrasında yorumlanmıştır. Bu parametreler pencere uzunluğu, üst üste binme oranı ve FFT sayısı parametreleridir. Bu üç adet parametre ayrı ayrı incelenerek etkileri deneysel olarak gösterilmiştir. Böylelikle spektral tepeler yardımıyla ses parmak izi temelli yapılacak çalışmalarda, incelenen bu yöntemler ile bu alana katkı sağlanması hedeflenmiştir.

Uygulanan tüm testlerde algoritma bu işlemi büyük bir başarı ile gerçekleştirmiştir. Yapılan testler sonucunda, tanıma süresi, eşleşme sayısı ve gürültüye karşı sağlamlık birlikte dikkate alındığında, 512;512 spektrogram parametre kombinasyonunun en iyi sonuçları verdiği görülmüştür.

Zaman ve frekans eksenlerindeki değişimlere karşı gürbüzlük elde edilmesi ve özel tasarlanacak bazı filtreler ile program performansının iyileştirilmesi, ileriye dönük çalışmalar arasındadır.

KAYNAKLAR

- [1] **P. Cano, E. Batle, T. Kalker and J. Haitsma** (2002). "A review of algorithms for audio fingerprinting", *2002 IEEE Workshop on Multimedia Signal Processing*, pp. 169-173, doi: 10.1109/MMSP.2002.1203274.
- [2] **J. Haitsma and T. Kalker** (2002). A Highly Robust Audio Fingerprinting System. *Journal of New Music Research* 32 (2), pp. 211 – 221.
- [3] **A. Wang** (2003). "An Industrial-Strength Audio Search Algorithm". In: *Proc. International Society for Music Information Retrieval (ISMIR)*. pp. 7–13.
- [4] **S. Baluja and M. Covell** (2007). "Audio Fingerprinting: Combining Computer Vision & Data Stream Processing," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '07*, pp. II-213-II-216, doi: 10.1109/ICASSP.2007.366210.
- [5] **X. Anguera, A. Garzon and T. Adamek** (2012). "MASK: Robust Local Features for Audio Fingerprinting", *2012 IEEE International Conference on Multimedia and Expo*, pp. 455-460, doi: 10.1109/ICME.2012.137.
- [6] **A. Garzon** (2011). *Audio Fingerprinting* (Master's thesis). Retrieved from <http://mtg.upf.edu/node/2326>
- [7] **Fulop S.A.** (2011) History of Speech Spectrum Analysis. In: *Speech Spectrum Analysis. Signals and Communication Technology*. pp. 41-68. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17478-0_3
- [8] **J.B.J. Fourier** (1807). *Théorie de la propagation de la chaleur dans les solides*. Manuscript first published in I. Grattan-Guinness, Joseph Fourier 1768–1830 (The MIT Press, Cambridge, 1972).
- [9] **M.T. Heideman, D.H. Johnson, C.S. Burrus** (1984), Gauss and the history of the fast Fourier transform. *IEEE Acoust. Speech Sig. Proc. Mag.* 1, pp. 14–21.
- [10] **Juang, B. & Rabiner, Lawrence** (2005). *Automatic Speech Recognition-A Brief History of the Technology Development*.
- [11] **K. H. Davis, R. Biddulph, and S. Balashek**, (1952). Automatic Recognition of Spoken Digits, *J. Acoust. Soc. Am.*, Vol 24, No. 6, pp. 627-642.
- [12] **Abreha, G. T.** (2014). *An Environmental Audio-Based Context Recognition System Using Smartphones*, (Master's Thesis). Accessed on Jun 27, 2020, Retrieved from http://essay.utwente.nl/66444/1/Abreha_MA_EWI.pdf
- [13] **Hideaki Kanai, Toyohisa Nakada, Yusuke Hanbat and Susumu Kunifuji** (2008). "A support system for context awareness in a group home using sound cues," *2008 Second International Conference on Pervasive Computing Technologies for Healthcare*, pp. 264-267, doi: 10.1109/PCTHEALTH.2008.4571085.
- [14] **D. M. Lee, H. C. Lee, G. H. Nam, T. Kim and C. Jeong** (2020). "A Sound Source-based Intelligent Context Awareness System using CNN," *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 594-597, doi: 10.1109/ICAIIIC48513.2020.9065195.

- [15] **J. E. Rougui, D. Istrate and W. Soudene** (2009). "Audio sound event identification for distress situations and context awareness," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3501-3504, doi: 10.1109/IEMBS.2009.5334581.
- [16] **L. Huang and C. Wang** (2020). "WalkID: Towards Context Awareness of Smart Home by Identifying Walking Sounds," *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pp. 1-6, doi: 10.1109/WF-IoT48130.2020.9221184.
- [17] **M. Mielke and R. Brück** (2016). "AUDIS wear: A smartwatch based assistive device for ubiquitous awareness of environmental sounds," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5343-5347, doi: 10.1109/EMBC.2016.7591934.
- [18] **B. V. Varun, S. M. Kusuma and G. REDDY** (2020). "AI-edge based voice responsive smart headphone for user context-awareness," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 1-5, doi: 10.1109/CONECCT50063.2020.9198484.
- [19] **S. Chu, S. Narayanan, C., J. Kuo and M. J. Mataric** (2006). "Where am I? Scene Recognition for Mobile Robots using Audio Features," *2006 IEEE International Conference on Multimedia and Expo*, pp. 885-888, doi: 10.1109/ICME.2006.262661.
- [20] **Jie Huang** (2002). "Spatial auditory processing for a hearing robot," *Proceedings IEEE International Conference on Multimedia and Expo*, 2002, pp. 253-256 vol.2, doi: 10.1109/ICME.2002.1035566.
- [21] **C. Clavel, T. Ehrette and G. Richard** (2005). "Events Detection for an Audio-Based Surveillance System," *2005 IEEE International Conference on Multimedia and Expo*, pp. 1306-1309, doi: 10.1109/ICME.2005.1521669.
- [22] **J. Gonzalez-Rodriguez, J. Fierrez-Aguilar and J. Ortega-Garcia** (2003). "Forensic identification reporting using automatic speaker recognition systems," *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. (ICASSP '03).*, 2003, pp. II-93, doi: 10.1109/ICASSP.2003.1202302.
- [23] **R. Dong, D. Hermann, E. Cornu and E. Chau** (2007). "Low-power implementation of an HMM-based sound environment classification algorithm for hearing aid application," *2007 15th European Signal Processing Conference*, pp. 1635-1638.
- [24] **Chen J., Kam A.H., Zhang J., Liu N., Shue L.** (2005) Bathroom Activity Monitoring Based on Sound. In: *Gellersen H.W., Want R., Schmidt A. (eds) Pervasive Computing. Pervasive 2005. Lecture Notes in Computer Science, vol 3468.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/11428572_4
- [25] **M. Vacher, F. Portet, A. Fleury and N. Noury** (2010). "Challenges in the processing of audio channels for Ambient Assisted Living," *The 12th IEEE International Conference on e-Health Networking, Applications and Services*, 2010, pp. 330-337, doi: 10.1109/HEALTH.2010.5556546.

- [26] **J. Wang, H. Lee, J. Wang and C. Lin** (2008). "Robust Environmental Sound Recognition for Home Automation," in *IEEE Transactions on Automation Science and Engineering*, vol. 5, no. 1, pp. 25-31, doi: 10.1109/TASE.2007.911680.
- [27] **M. Muller, D. P. W. Ellis, A. Klapuri and G. Richard** (2011). "Signal Processing for Music Analysis," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1088-1110, doi:10.1109/JSTSP.2011.2112333.
- [28] **C. Roads** (1996). *The Computer Music Tutorial*. Cambridge, MA: MIT Press.
- [29] **M. S. Puckette** (2007). *The Theory and Technique of Electronic Music*. Singapore: World Scientific.
- [30] **Chroma Feature Analysis and Synthesis**, Erişim tarihi: 22.01.2021 <https://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/>.
- [31] **D. P. W. Ellis and G. E. Poliner** (2007). "Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP '07*, pp. IV-1429-IV-1432, doi: 10.1109/ICASSP.2007.367348.
- [32] **S. Ewert, M. Muller and P. Grosche** (2009). "High resolution audio synchronization using chroma onset features," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1869-1872, doi: 10.1109/ICASSP.2009.4959972.
- [33] **J. Serra, E. Gomez, P. Herrera and X. Serra** (2008). "Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6, pp. 1138-1151, doi: 10.1109/TASL.2008.924595.
- [34] **T. Bertin-Mahieux, R. J. Weiss, and D. P. W. Ellis** (2010). "Clustering beat-chroma patterns in a large music database," in *Proc. Int. Conf. Music Inf. Retrieval (ISMIR)*, Utrecht, The Netherlands, pp. 111–116.
- [35] **M. A. Bartsch and G. H. Wakefield** (2001). "To catch a chorus: using chroma-based representations for audio thumbnailing," *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 15-18, doi: 10.1109/ASPAA.2001.969531.
- [36] **Parncutt, R.** (1994). A Perceptual Model of Pulse Saliency and Metrical Accent in Musical Rhythms. *Music Perception: An Interdisciplinary Journal*, 11(4), 409-464. doi:10.2307/40285633
- [37] **Chord**, Britannica. Erişim tarihi: 03.03.2021. Encyclopedia Britannica, <https://www.britannica.com/art/chord-music>
- [38] **Fujishima, T.** (1999). "Realtime chord recognition of musical sound: A system using common lisp music," in *Proc. ICMC*, Beijing, China, pp. 464–467.
- [39] **M. Mauch and S. Dixon** (2010). "Approximate note transcription for the improved identification of difficult chords," in *Proc. 11th Int. Soc. Music Inf. Retrieval Conf. (ISMIR)*, Utrecht, The Netherlands, pp. 135–140.

- [40] **A. Degani, M. Dalai, R. Leonardi and P. Migliorati** (2015). "Harmonic Change Detection for musical chords segmentation," *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177404.
- [41] **I. Rida, R. Herault and G. Gasso** (2014). "Supervised Music Chord Recognition," *2014 13th International Conference on Machine Learning and Applications*, pp. 336-341, doi: 10.1109/ICMLA.2014.60.
- [42] **L. Oudre, Y. Grenier and C. Fevotte** (2011). "Chord Recognition by Fitting Rescaled Chroma Vectors to Chord Templates," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2222-2233, doi: 10.1109/TASL.2011.2139205.
- [43] **T. Gagnon, S. Larouche and R. Lefebvre** (2003). "A neural network approach for preclassification in musical chords recognition," *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, pp. 2106-2109 Vol.2, doi: 10.1109/ACSSC.2003.1292351.
- [44] **B. Lin and T. Yeh** (2017). "Automatic Chord Arrangement with Key Detection for Monophonic Music," *2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIT)*, pp. 21-25, doi: 10.1109/ICSIT.2017.29.
- [45] **V. Zenz and A. Rauber** (2007). "Automatic Chord Detection Incorporating Beat and Key Detection," *2007 IEEE International Conference on Signal Processing and Communications*, pp. 1175-1178, doi: 10.1109/ICSPC.2007.4728534.
- [46] **M. Mauch and S. Dixon** (2010). "Simultaneous estimation of chords and musical context from audio," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1280–1289, Aug. 2010.
- [47] **A. Shehand, D.P.W. Ellis** (2003). "Chord segmentation and recognition using EM-trained hidden Markov models," in *Proc. Int. Conf. Music Inf. Retrieval (ISMIR)*, Baltimore, MD.
- [48] **J. P. Bello and J. Pickens** (2005). "A robust mid-level representation for harmonic content in music signals," in *Proc. Int. Conf. Music Inf. Retrieval (ISMIR)*, London, U.K.
- [49] **K. Lee and M. Slaney** (2008). "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 2, pp. 291–301.
- [50] **H. Papadopoulos and G. Peeters** (2008). "Simultaneous estimation of chord progression and downbeats from an audio file," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 121–124.
- [51] **M. Mauch, K. Noland, and S. Dixon** (2009). "Using musical structure to enhance automatic chord transcription", in *Proc. Int. Conf. Music Inf. Retrieval (ISMIR)*, Kobe, Japan, pp. 231–236.
- [52] **Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono, and S. Sagayama** (2010). "HMM-based approach for automatic chord detection using refined acoustic features," in *Proc. 35nd IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, pp. 5518–5521.

- [53] **Christopher Harte, Mark S, & Samer Abdallah** (2005). Symbolic Representation Of Musical Chords: A Proposed Syntax For Text Annotations. In *Proc. 4th Int. Conf. on Music Information Retrieval (ISMIR)*, 2005 (pp. 66–71).
- [54] **Daniel Jurafsky and H James** (2000). *Speech and language processing an introduction to natural language processing, computational linguistics, and speech*. Online available. Retrived date: 22.03.2021. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- [55] **Tan, L., & Jiang, J.** (2019). *Signal Sampling and Quantization*. Digital Signal Processing, pp. 13–58. Academic Press.isbn: 9780128150719.
- [56] **Paul R. Hill** (2019), *Audio and Speech Processing With MATLAB*. Frequency Analysis for Audio, pp. 56-97.
- [57] **Giannakopoulos, T., & Pikrakis, A.** (2014). *Signal Transforms and Filtering Essentials*. Introduction to Audio Analysis, pp. 33–57.
- [58] **Cookbook formulae for audio EQ biquad filter coefficients**. Eriřim tarihi: 22.02.2021. http://www.musicdsp.org/files/Audio-EQ_Cookbook.txt
- [59] **Andreas Antoniou** (2006). *Digital signal processing*. McGraw-Hill Toronto, Canada.
- [60] **A. Oppenheim, R. Schafer, and J. Buck** (1999), *Discrete-Time Signal Processing*, second edition, Prentice-Hall.
- [61] **Van Trees** (2002). H.L.: ‘*Optimum array processing*’ (Wiley Interscience, New York, NY, USA.
- [62] **Doerry, A.W.** (2017). ‘*Catalog of window taper functions for sidelobe control*’. Sandia National Laboratories, Albuquerque, New Mexico, USA.
- [63] **Prabhu, K.** (2014). *Window Functions And Their Applications In Signal Processing*. Boca Raton [Florida]: CRC Press/Taylor & Francis.
- [64] **F. J. Harris (1978)**. "On the use of windows for harmonic analysis with the discrete Fourier transform," in *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51-83, doi: 10.1109/PROC.1978.10837.
- [65] **Gade, Svend, and Henrik Herlufsen** (1987). “*Use of Weighting Functions in DFT/FFT Analysis (Part I)*.” Windows to FFT Analysis (Part I): Brüel & Kjaer Technical Review. Vol. X, pp. 1–28.
- [66] **B.R. Mahafza** (2005), *Radar Systems Analysis and Design Using MATLAB*, CRC Press, 2nd edn., Florida, USA.
- [67] **S. Qian and D. Chen** (1993). "Discrete Gabor transform," in *IEEE Transactions on Signal Processing*, vol. 41, no. 7, pp. 2429-2438, doi: 10.1109/78.224251.
- [68] **E. Jacobsen and R. Lyons** (2003). "The sliding DFT," in *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 74-80, doi: 10.1109/MSP.2003.1184347.
- [69] **G. Schmalz** (2012). Encoding images as sound & decoding via spectrogram, [Online]. Available:<http://www.gramschmalz.com/encoding-images-as-sounddecoding-via-spectrogram/>
- [70] **J. O. Smith. III** (2007). *Mathematics of the Discrete Fourier Transform (DFT), with Audio Applications* — Second Edition. W3K Publishing.

- [71] **L. Rabiner and J. Allen** (1977). "A unified approach to short-time fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. N/A, pp. 1558–1564.
- [72] **M. Bosi and R. E. Goldberg** (2003). *Introduction to Digital Audio Coding and Standards*. Kluwer Academic Publishers, isbn: 978-1-4613-5022-4.
- [73] **M. Müller** (2007). *Information Retrieval for Music and Motion*. Springer, 2007.
- [74] **K. R. Rao, D. N. Kim, and J. J. Hwang** (2011). *Fast Fourier Transform - Algorithms and Applications*. Springer Science & Business Media.
- [75] **J. S. Walker and G. W. Don** (2013). *Mathematics and Music: Composition, Perception, and Performance*. CRC Press, 2013. pp. 93-129.
- [76] **Practical Introduction to Time-Frequency Analysis**, Erişim tarihi: 05.01.2021. <https://www.mathworks.com/help/signal/examples/practical-introduction-to-time-frequency-analysis.html>
- [77] **Kim Hyoung-Gook, Moreau Nicolas, Thomas Sikora** (2005). *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, JohnWiley & Sons.
- [78] **S. Baluja and M. Covell**. (2008). "Waveprint: Efficient wavelet-based audio fingerprinting". In: *Pattern Recognition 41.11*. pp. 3467–3480.
- [79] **J. Haitsma, T. Kalker, and J. Oostveen** (2001). "Robust audio hashing for content identification". In: *International Workshop on Content-Based Multimedia Indexing*. Vol. 4, pp. 117–124.
- [80] **E. Younessian et al.** (2010). "Telefonica Research at TRECVID 2010 Content-Based Copy Detection". In: *Proc. of TRECVID. 2010*.
- [81] **C. Herley** (2006). "ARGOS: automatically extracting repeating objects from multimedia streams," in *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 115-129, doi: 10.1109/TMM.2005.861286.
- [82] **B. Coover and J. Han** (2014). "A Power Mask based audio fingerprint," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1394-1398, doi: 10.1109/ICASSP.2014.6853826.
- [83] **Tsai T.** (2016). *Audio Hashprints: Theory & Application*. (Master Thesis). Retrived from: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-185.pdf>
- [84] **C. C. Liu, J. L. Hsu, & A. L. P. Chen** (1999). An approximate string matching algorithm for content-based music data retrieval. *IEEE International Conference on Multimedia Computing and Systems*, pages 451–456.
- [85] **M. Malekesmaeili and R. K. Ward** (2012), "A novel local audio fingerprinting algorithm," *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 136-140, doi: 10.1109/MMSP.2012.6343429.
- [87] **D.G. Lowe** (2004). Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110.
- [88] **Pedro Cano, Eloi Batlle, Harald Mayer, and Helmut Neuschmied** (2002). Robust sound modeling for song detection in broadcast audio. In *Proceedings of the 112th AES Convention*, pages 1–7.

- [89] **A. Ramalingam and S. Krishnan** (2005). "Gaussian mixture modeling using short time fourier transform features for audio fingerprinting". In: *IEEE International Conference on Multimedia and Expo. 2005*, pp. 1146–1149.
- [90] **C.W. Ngo et al.** (2009). "VIREO/DVMM at TRECVID 2009: High-Level Feature Extraction, Automatic Video Search, and Content-Based Copy Detection". In: *Proc. of TRECVID*. pp. 415–432.
- [91] **Y. Liu et al.** (2010). "Coherent Bag-Of Audio Words Model for Efficient Large-Scale Video Copy Detection". In: *Proc. ACM International Conference on Image and Video Retrieval*. pp. 89–96.
- [92] **C.C. Liu and P.F. Chang** (2011). "An Efficient Audio Fingerprint Design for MP3 Music". In: *Proc. ACM International Conference on Advances in Mobile Computing and Multimedia (MoMM'11)*. pp. 190–193.
- [93] **H. Khemiri, D. Petrovska-Delacrétaz and G. Chollet** (2014), "Detection of repeating items in audio streams using data-driven ALISP sequencing," *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 446-451, doi: 10.1109/ATSIP.2014.6834653.
- [94] **H. Jégou et al.** (2012). "Babaz: a large scale audio search system for video copy detection". In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 2369–2372.
- [95] **Jaap Haitsma and Ton Kalker** (2003). Speed-change resistant audio fingerprinting using autocorrelation. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 728–731.
- [96] **Eric Allamanche, Jürgen Herre, Oliver Hellmuth, Bernhard Fröba, Thorsten Kastner, and Markus Cremer** (2001). Content-based identification of audio material using mpeg-7 low level description. In *ISMIR 2001*.
- [97] **Eric Allamanche, Jürgen Herre, Oliver Hellmuth, Bernhard Fröba, and Markus Cremer** (2001). AudioID: Towards content-based identification of audio material. In *Proceedings of the 110th AES Convention*, Amsterdam, NL.
- [98] **M. Ramona and G. Peeters** (2011). "Audio identification based on spectral modeling of bark-bands energy and synchronization through onset detection," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 477-480, doi: 10.1109/ICASSP.2011.5946444.
- [99] **Sébastien Fenet, Gaël Richard, and Yves Grenier** (2011). A scalable audio fingerprint method with robustness to pitch-shifting. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, Miami, USA.
- [100] **E. Dupraz and G. Richard** (2010). "Robust frequency-based Audio Fingerprinting," *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 281-284, doi: 10.1109/ICASSP.2010.5495944.
- [101] **J. Six and M. Leman** (2014). "Panako: a scalable acoustic fingerprinting system handling time-scale and pitch modification". In: *Proc. International Society for Music Information Retrieval (ISMIR)*.
- [102] **R. Sonnleitner and G. Widmer** (2014). "Quad-based audio fingerprinting robust to time and frequency scaling". In: *Proc. International Conference on Digital Audio Effects*. pp. 173–180.

- [103] **J. George and A. Jhunjhunwala** (2015), "Scalable and robust audio fingerprinting method tolerable to time-stretching," *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 436-440, doi: 10.1109/ICDSP.2015.7251909.
- [104] **Yan Ke, D. Hoiem and R. Sukthankar** (2005), "Computer vision for music identification," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 597-604 vol. 1, doi: 10.1109/CVPR.2005.105.
- [105] **J. S. Seo, Minho Jin, Sunil Lee, Dalwon Jang, Seungjae Lee and C. D. Yoo** (2005). "Audio fingerprinting based on normalized spectral subband centroids," *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. iii/213-iii/216 Vol. 3, doi: 10.1109/ICASSP.2005.1415684.
- [106] **J. S. Seo, Minho Jin, Sunil Lee, Dalwon Jang, Seungjae Lee and C. D. Yoo** (2006). "Audio fingerprinting based on normalized spectral subband moments," in *IEEE Signal Processing Letters*, vol. 13, no. 4, pp. 209-212, doi: 10.1109/LSP.2005.863678.
- [107] **Park, M., Kim, H.-R. and Yang, S.H.** (2006), Frequency-Temporal Filtering for a Robust Audio Fingerprinting Scheme in Real-Noise Environments. *ETRI Journal*, 28: 509-512. <https://doi.org/10.4218/etrij.06.0205.0135>
- [108] **S. Kim and C. D. Yoo (2007)**. "Boosted Binary Audio Fingerprint Based on Spectral Subband Moments," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '07*, pp. I-241-I-244, doi: 10.1109/ICASSP.2007.366661.
- [109] **J. Lebossé, L. Brun, and J.C. Pailles (2007)**. "A robust audio fingerprint extraction algorithm". In: *Proceedings of IASTED International Conference: Signal Processing, Pattern Recognition, and Applications*. pp. 269–274.
- [110] **A. Saracoglu et al.** (2009). "Content Based Copy Detection with Coarse Audio-Visual Fingerprints," *2009 Seventh International Workshop on Content-Based Multimedia Indexing, 2009*, pp. 213-218, doi: 10.1109/CBML.2009.12.
- [111] **R. Mukai et al.** (2010). "NTT Communication Science Laboratories at TRECVID 2010 Content Based Copy Detection". In: *Proc. of TRECVID*.
- [112] **St Denis, T., & Johnson, S.** (2007). Hash Functions. *Cryptography for Developers*, pp. 203–250.
- [113] **Cano, P., Batlle, E., Kalker, T. et al.** (2005). A Review of Audio Fingerprinting. *J VLSI Sign Process Syst Sign Image Video Technol* 41, pp. 271–284. <https://doi.org/10.1007/s11265-005-4151-3>
- [114] **L. Boney, A.I. Tewfik, and K.I. Hamdy** (1996). Digital Watermarks for Audio Signals, *Proceedings EUSIPCO-96, Vol. III, VIII European Signal Processing Conference; Trieste, Italy*, S. 473–480.
- [115] **Webster, J.G. and Steinebach, M.** (2015). *Digital Audio Watermarking*. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, J.G. Webster (Ed.). <https://doi.org/10.1002/047134608X.W8281>

- [116] **Cano P., Batlle E., Gómez E., de C.T. Gomes L., Bonnet M.** (2005) Audio Fingerprinting: Concepts And Applications. In: *K. Halgamuge S., Wang L. (eds) Computational Intelligence for Modelling and Prediction. Studies in Computational Intelligence*, vol 2. Springer, Berlin, Heidelberg. https://doi.org/10.1007/10966518_17
- [117] **Gerhard Heinzl, Albrecht Rüdiger, and Roland Schilling** (2002). Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows. Erişim Tarihi: 02.26.2020, https://holometer.fnal.gov/GH_FFT.pdf
- [118] **Nasir Ahmed, T. Natarajan, and Kamisetty R. Rao** (1974). Discrete cosine transform. *IEEE transactions on Computers*, 100(1): pp. 90-93.
- [119] **S. R. Subramanya, R. Simha, B. Narahari and A. Youssef**, "Transform-based indexing of audio data for multimedia databases," *Proceedings of IEEE International Conference on Multimedia Computing and Systems, 1997*, pp. 211-218, doi: 10.1109/MMCS.1997.609595.
- [120] **Mihçak M.K., Venkatesan R.** (2001) A Perceptual Audio Hashing Algorithm: A Tool for Robust Audio Identification and Information Hiding. In: *Moskowitz I.S. (eds) Information Hiding. IH 2001. Lecture Notes in Computer Science*, vol 2137. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45496-9_5
- [121] **S. Theodoridis and K. Koutroumbas** (1999). *Pattern recognition*. Academic press. New York.
- [122] **Constantin Papaodysseus, George Roussopoulos, Dimitrios Fragoulis, Athanasios Panagopoulos, and Constantin Alexiou** (2001). *A new approach to the automatic recognition of musical recordings*. Journal of the Audio Engineering Society, 49(1/2):23-35.
- [123] **Joseph W Picone** (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9):1215-1247.
- [124] **Request for Information on Audio Fingerprinting Technologies** (2001). Retrieved from: [http://www.riaa.org/pdf/RIAA IFPI Fingerprinting RFI.pdf](http://www.riaa.org/pdf/RIAA_IFPI_Fingerprinting_RFI.pdf). 2001.
- [125] **P. Cano, M. Kaltenbrunner, O. Mayor, and E. Batlle** (2001). "Statistical Significance in Song-Spotting in Audio," *Proceedings of the International Symposium on Music Information Retrieval*, Bloomington, IN.
- [126] **Shazam data is powering Apple Music's newest chart, the Shazam Discovery Top 50**, Accessed on: May 30, 2020, Retrived from: <https://techcrunch.com/2019/08/20/shazam-data-is-powering-apple-musics-newest-chart-the-shazam-discovery-top-50/>
- [127] **Scerri, E.** (2016). "*Music recognition, analysis, and automatic tagging using spectrogram pattern matching.*" Retrieved from: <http://www.cs.rug.nl/~george/wp-content/uploads/2016/08/ScerriErik.pdf>
- [128] **H. Zhang, G. Hua, L. Yu, Y. Cai, G. Bi** (2017). Underdetermined blind separation of overlapped speech mixtures in time–frequency domain with estimated number of sources, *Speech Commun.* Volume 89, Pages 1-16, ISSN 0167-6393, <https://doi.org/10.1016/j.specom.2017.02.003>.

- [129] **Shaik, B. S., Naganjaneyulu, G. V. S. S. K. R., Chandrasheker, T., & Narasimhadhan, A. V.** (2015). A Method for QRS Delineation Based on STFT Using Adaptive Threshold. *Procedia Computer Science*, 54, 646–653.
- [130] **Robust Landmark-Based Audio Fingerprinting** (2009), D. Ellis, web resource, available: <http://labrosa.ee.columbia.edu/matlab/fingerprint/>
- [131] **N. Liu, J. Gao, X. Jiang, Z. Zhang and Q. Wang** (2017). "Seismic Time–Frequency Analysis via STFT-Based Concentration of Frequency and Time," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 127–131.
- [132] **R. Bracewell** (1986). *The Fourier Transform and Its Applications*, McGraw-Hill.
- [133] **L. Cohen** (1995). *Time-Frequency Analysis*, Electrical Engineering Signal Processing, Prentice Hall PTR.
- [134] **S. Qian and Dapang Chen** (1999). Joint Time-Frequency Analysis, *IEEE Signal Processing Magazine*, vol. 16(2), pp. 52–67.
- [135] **D. Gabor** (1946). "Theory of communication. Part 1: the analysis of information," *Journal of the Institution of Electrical Engineers—Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441.
- [136] **Castanié Francis** (1998). *Digital spectral analysis: parametric, non-parametric and advanced methods*. London: ISTE, 2011. p.290.
- [137] **Boashash B.** (2003). *Heuristic Formulation of Time-Frequency Distributions*. Signal Processing Research Centre, Queensland University of Technology, Brisbane, Australia. Time Frequency Analysis, pp. 29–57, doi: 10.1016/B978-008044335-5/50023-1
- [138] **Werner Heisenberg** (1985). *Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik*. In *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pages 478–504. Springer.
- [139] **Practical Introduction to Time-Frequency Analysis, Trading Off Time and Frequency Resolution to Get the Best Representation of Your Signal.** Accessed: May 29, 2020. [Online]. Available: <https://www.mathworks.com/help/signal/examples/practicalintroduction-to-time-frequency-analysis.html>
- [140] **K. K. Paliwal, J. G. Lyons and K. K. Wójcicki** (2010). "Preference for 20-40 ms window duration in speech analysis," *2010 4th International Conference on Signal Processing and Communication Systems*, pp. 1-4, doi: 10.1109/ICSPCS.2010.5709770.
- [141] **P. Cuff**, ELE301: Signals and Systems-Labs., Princeton University, Fall Semester 2011-12, Lab. 5. Accessed on: May 30, 2020. [Online]. Available: https://www.princeton.edu/~cuff/ele301/files/Lab6_2011.pdf.
- [142] **Grosche, P., Müller, M., & Serrà, J.** (2012). *Audio Content-Based Music Retrieval*. Multimodal Music Processing. Dagstuhl Follow-Ups, Vol. 3. ISBN 978-3-939897-37-8. pp. 157–174
- [143] **McClaning, K. and Vito, T.** (2000). *Radio Receiver Design*. Atlanta, GA: Noble Pub. Corp.

ÖZGEÇMİŞ

ÖĞRENİM DURUMU:

- **Lisans** : 2016, İnönü Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü
- **Yüksek Lisans** : 2021, İnönü Üniversitesi, Elektrik Elektronik Mühendisliği Anabilim Dalı

MESLEKİ DENEYİM:

- 2018-2021 Munzur Üniversitesi'nde çalıştı.

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN ÇALIŞMALAR

- **Koseoglu, M., & Uyanik, H. (2020, June).** The Effect of Different Noise Levels on The Performance of The Audio Search Algorithm. 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA-IEEE). <https://doi.org/10.1109/hora49412.2020.9152913>.
- **Uyanik, H., & Koseoglu, M. (2020, October).** Performance Evaluation of Different Window Functions for Audio Fingerprint Based Audio Search Algorithm. 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT-IEEE). <https://doi.org/10.1109/ismsit50672.2020.9255225>.